Final Exam

Multilevel Models (Autumn 2019)

MAE4112 Multilevel Models
Centre for Educational Measurement at the University of Oslo (CEMO)

Date and time: 13 December 2019, 09:00-13:00

Welcome to the MAE4112 Multilevel Models exam!

During the next **4 hours**, you will be working on **33 tasks** which are structured in 8 units. In total, you will be able to get 86 credits.

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

1. **Read** the questions **carefully**.
2. Notice which **task operators** are used (e.g., *name* something vs. *describe* something).
3. You may **simplify subscripts** wherever appropriate (e.g., Yij instead of Y_{ij}).
4. You may also **simplify Greek letters** wherever appropriate (e.g., beta0j instead of β_{0j}, gamma00 instead of γ_{00}).
5. Use **dots instead of commas** to indicate decimals.
6. Keep your **explanations and descriptions brief**.
7. **Partial credits** will be given.

Some abbreviations you may want to use throughout the exam:
L1—Level 1, within level
L2—Level 2, between level
ICC—Intraclass correlation

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

Best regards,
Chia-Wen and Ronny

Name:	SUGGESTED SOLUTIONS + GRADING

Results

Task	Credits	Max. credits
C1		9
C2		4
A1		8
A2		16
A3		19
D1		14
D2		10
D3		6
TOTAL:		**86**

Time estimation

Time needed to complete the exam [instructors]: 70 minutes
Time needed to complete the exam, estimated for students: 70*3=210 minutes

Grading

Grade	Credit intervals	Percentages correct
A	86-77	100-90 %
B	76-69	89-80 %
C	68-60	79-70 %
D	59-52	69-60 %
E	51-43	59-50 %
F	42-0	49-0 %

C1. Key Concepts in Multilevel Modeling (9 credits)

Definition of terms and concepts

This section focuses on the key concepts and terms used in multilevel modeling. You will be asked to provide some definitions, explanations, and examples.

a) Explain the following terms and concepts briefly.

Term or concept	Explanation
Climate variable	A climate variable refers to a L2 variable which was aggregated from a L1 variable; the L1 variable represents ratings/perceptions of the L2 units. Example: classroom-average ratings of instruction.
Intraclass correlation ICC_2	The ICC_2 represents the reliability of the group means. It is specific to each group, but can be averaged to an overall value. $$Rel(\bar{X}_{.j}) = ICC_{2j} = \lambda_j = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_r / n_j}$$
Grand mean centering	A form of centering L1 or L2 predictors by subtracting the overall (grand) mean of a variable from the original variable, $$X^C_{ij} = X_{ij} - \bar{X}_{..} \text{ or } Z^C_j = Z_j - \bar{Z}_{.}$$
Cross-level interaction effect	This represents the effect of a L2 variable on the relation between two L1 variables. It also represents the regression coefficient of an interaction variable that is the product of a L1 and a L2 predictor variable on the L1 outcome variable (after controlling for the L1 and L2 predictor variables).

SCORING:

- 1 credit per correct explanation or definition
- **Subtotal: 4 credits**

b) Provide the L1 and L2 equations for a two-level random-slope model.

Note: There is no need to specify the distributions and structure of the random effects.

Random-Slope Model
Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$ Level 2: $$\beta_{0j} = \gamma_{00} + u_{0j}$$ $$\beta_{1j} = \gamma_{10} + u_{1j}$$

SCORING:

- 1 credit per correct equation
- **Subtotal: 3 credits**

c) A researcher specifies a two-level random-slope model. What could be a possible research question underlying this model?

Note: You may provide a specific example or describe the type of research questions in general. In the case of a specific example, please do not forget to name the levels L1 and L2.

Research question
For instance, To what extent does the relation between students' SES and mathematics achievement vary between schools? General: To what extent does the relation between two L1 variables vary between L2 units?

SCORING:

- 1 credit for mentioning the relation between two L1 variables
- 1 credit for the mentioning of variation in the slope at L2
- **Subtotal: 2 credits**

C2. Multilevel Mediation Models (4 credits)

The PISA 2015 study collected data from students (L1) who were nested in schools (L2). In the following, you will find a list of variables and their descriptions from this data set.

Variable	Description
SCHOOLID	School identifier (ID)
STUDID	Student identifier (ID)
FEMALE	Student Gender coded as 1=Female and 0=Male
PV1MATH	Student Achievement in Mathematics
PV1READ	Student Achievement in Reading
PV1SCIE	Student Achievement in Science
IMMIG	Student Immigration Status
AGE	Student Age
TEACHSUP	Student Perception of the Academic Support given by Teachers
JOYSCIE	Student Enjoyment of Science
MOED	Mother's Education
FAED	Father's Education
SFFT	Student rating to Safety of Facilities in School
BELONG	Student Sense of Belonging to School
ANXTEST	Student Test Anxiety
MOTIVAT	Student Achievement Motivation
UNFAIR	Student rating to Teacher Fairness
WEALTH	Family Wealth
ESCS	Index of Economic, Social, and Cultural Status
SCHSIZE	School Size
STRATIO	Student-Teacher ratio
PRIVATE	School Ownership
schPV1MATH	School-average Achievement in Mathematics
schPV1READ	School-average Achievement in Reading
schPV1SCIE	School-average Achievement in Science
schANXTEST	School-average Test Anxiety
schSFFT	School-average Safety of Facilities in School
schESCS	School-average ESCS
schTEACHSUP	School-average Academic Support given by Teachers
schUNFAIR	School-average Teacher Fairness

Suppose a researcher intends to specify multilevel mediation models in which X represents the independent variable, M the mediator, and Y the dependent variable.

a) Select one variable each for X, M, and Y from the list of variables that could fit to a **2-1-1** multilevel mediation design.

Element	Possible variable
X	SCHSIZE
M	ESCS
Y	ANXTEST

NOTE:

- The variable X must be measured at L2 only; school-average variables that were derived from L1 variables do not count.

SCORING:

- 1 credit for the correct selection
- **Subtotal: 1 credit**

b) Select one variable each for X, M, and Y from the list of variables above that could fit to a **1-2-2** multilevel mediation design.

Element	Possible variable
X	ESCS
M	STRATIO
Y	SCHSIZE

NOTE:

- The variables M and Y must be measured at L2 only; school-average variables that were derived from L1 variables do not count.

SCORING:

- 1 credit for the correct selection
- **Subtotal: 1 credit**

c) Suppose the researcher is interested in examining the contextual effects on students' individual achievement in mathematics ($PV1MATH$). Select **two possible contextual variables** from the table above.

Possible contextual variables
schESCS, SCHSIZE

SCORING:

- 1 credit for the correct selection of two contextual variables
- **Subtotal: 1 credit**

d) Suppose the researcher is interested in examining the effects of climate variables on students' individual achievement in reading ($PV1READ$). Select **two possible climate variables** from the table above.

Possible climate variables
schTEACHSUP, schSFFT

SCORING:

- 1 credit for the correct selection of two contextual variables
- **Subtotal: 1 credit**

A1. In the beginning, there was the Null Model (8 credits)

In the following, a two-level null model is specified for a dependent variable Y_{ij} (i: students, j: schools).

Null Model	
Level 1 (students):	$Y_{ij} = \beta_{0j} + r_{ij}$
Level 2 (schools):	$\beta_{0j} = \gamma_{00} + u_{0j}$

a) Describe briefly the following elements of this model.

Elements of the model	Description
β_{0j}	Group-mean of Y (i.e., the mean of the variable Y within a cluster/group j)
r_{ij}	Level-1 residual (i.e., the deviation of an individual i in group j from the overall/grand mean of Y)
γ_{00}	Overall/grand mean of the variable Y
u_{0j}	Group-specific deviation from the grand mean of Y; random-effect component for group j

SCORING:

- 1 credit per correct description
- **Subtotal: 4 credits**

b) Define the intraclass correlation $\rho_I = ICC_1$ on the basis of the null model.

Definition of the ICC_1
$\rho_I = ICC_1 = \sigma^2_{u_0} / (\sigma^2_{u_0} + \sigma^2_r)$ or $\rho_I = Corr(Y_{ij}, Y_{lj})$ for $i \neq l$ students in school j

NOTES:

- A verbal description of the ICC_1 can also be provided (e.g., ratio between the between-level variance of a variable Y and the sum of the between- and within-level variance)
- ρ_I is correlation between the variables measured for two randomly selected different individuals in the same group.

SCORING:

- 1 credit for one correct definition
- **Subtotal: 1 credit**

A researcher has worked on the PISA 2015 data and **specified the null model for the variable named "BELONG"**, that is, students' sense of belonging to school. The PISA data follow a two-level structure with students (L1) nested in schools (L2). To identify the school for each student, the variable "SCHOOLID" is used. In total, J schools were included in the analyses.

This researcher obtained the following R output using the package lme4:

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BELONG ~ 1 + (1 | SCHOOLID)
##    Data: pisa15
##
## REML criterion at convergence: 80263
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1959 -0.5778 -0.0924  0.4939  2.3977
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SCHOOLID (Intercept) 0.0161   0.1269
##  Residual             1.2406   1.1138
## Number of obs: 26189, groups:  SCHOOLID, 1046
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 1.266e-01  8.067e-03 9.813e+02    15.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c) Assign the estimates (i.e., their values) in the output to the elements of the null model.
Note: For numbers with decimals, please report 4 decimals.

Elements of the model	Estimate in the R output
γ_{00}	0.1266
σ_r^2	1.2406
$\sigma_{u_0}^2$	0.0161
J	1046

SCORING:

- 0.5 credit per correct assignment
- **Subtotal: 2 credits**

d) Given the R output, how would you compute the intraclass correlation ICC[1] of BELONG, $\rho_I = ICC_1$?

Note: You do not need to compute the actual value of the ICC_1.

Computation of the ICC_1
$\rho_I = ICC_1 = 0.0161/(0.0161 + 1.2406)$

SCORING:

- 1 credit for the correct computation
- **Subtotal: 1 credit**

A2. School Climate and Mathematics Achievement (16 credits)

The Programme for International Student Achievement (PISA) 2015 assessed mathematics achievement for a large sample of 15-year-old students in the Nordic countries. The data (pisa15) contain the following variables:

- **SCHOOLID**: Identification variable of the school
- **PV1MATH**: Individual students' mathematics achievement (test score)
- **BELONG**: Individual students' sense of belonging to school
- **STUBEHA**: Student behavior in the school (rated by the principal in the school)
- **DISCLISCI**: Individual students' perceptions of the disciplinary climate
- **TEACHSUP**: Individual students' perceptions of the teacher support
- **STRATIO**: Student-teacher ratio in the school
- **gmDISCLISCI**: Group mean of perceptions of the disciplinary climate
- **gmTEACHSUP**: Group mean of perceptions of the teacher support

All variables were positively coded so that higher scores indicate better achievement, higher sense of belonging, better student behavior, more positive disciplinary climate, more teacher support, and a larger student-teacher ratio.

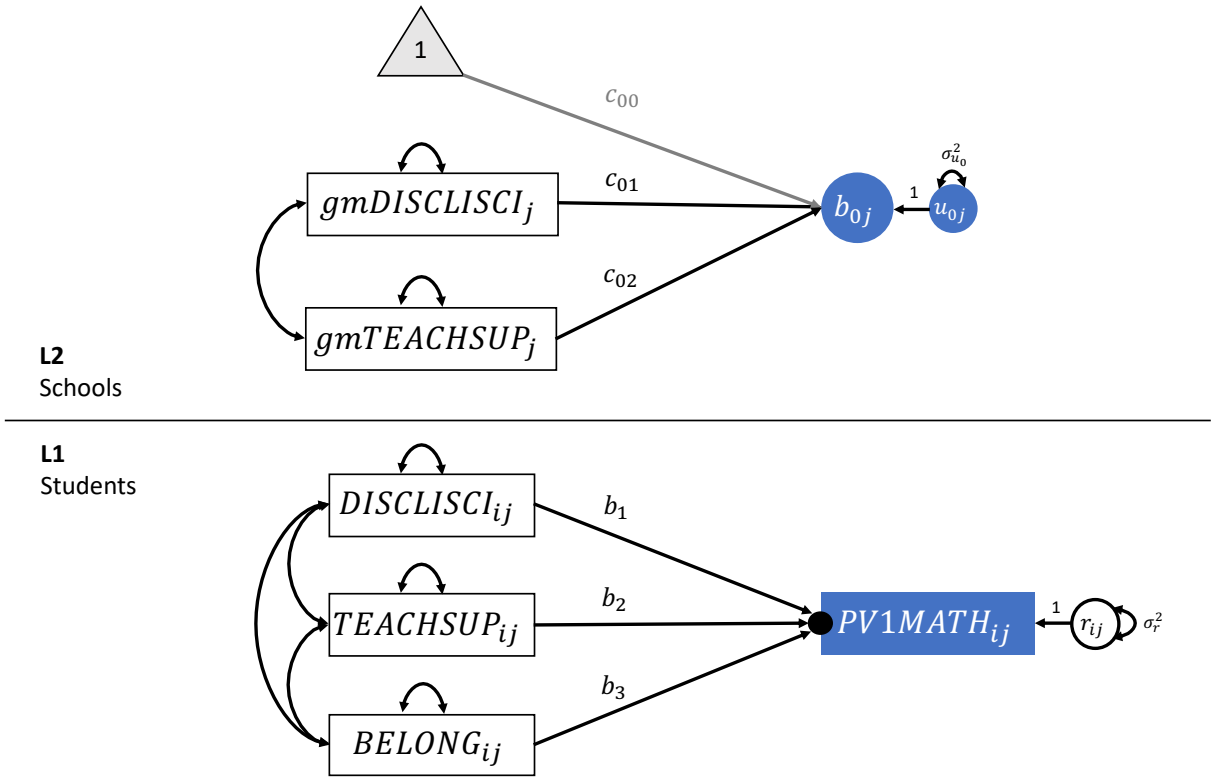
a) Determine whether the variables in the data set represent student- or school-level variables.

Variable	Student level	School level
PV1MATH	X	☐
BELONG	X	☐
STUBEHA	☐	X
DISCLISCI	X	☐
TEACHSUP	X	☐
STRATIO	☐	X
gmDISCLISCI	☐	X
gmTEACHSUP	☐	X

SCORING:

- 0.25 credits per correct variable
- **Subtotal: 2 credits**

Analyzing the data of *N* = 24775 students in 1040 schools, a researcher specified the following multilevel model:



b) Formulate the L1 and L2 model equations underlying this model.

Model specification
L1 (students): $$Y_{ij} = b_{0j} + b_{1j}DISCLISCI_{ij} + b_{2j}TEACHSUP_{ij} + b_{3j}BELONG_{ij} + r_{ij}$$ L2 (schools): $$b_{0j} = c_{00} + c_{01}gmDISCLISCI_j + c_{02}gmTEACHSUP_j + u_{0j}$$ $$b_{1j} = b_1$$ $$b_{2j} = b_2$$ $$b_{3j} = b_3$$

SCORING:

- 1 credit for the correct setup of the L1 equation (i.e., predictors, outcome)
- 1 credit for the correct setup of the L2 equation (i.e., predictors, outcome)
- 1 credit for the correct indices of variables at L1 and L2
- 1 credit for the acknowledgement of fixed slopes
- **Subtotal: 4 credits**

Estimating this model in lme4, the researchers obtained the following output:

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: PV1MATH ~ 1 + BELONG + DISCLISCI + TEACHSUP + gmDISCLISCI + gmTEACHSUP +
    (1 | SCHOOLID)
   Data: pisa15

REML criterion at convergence: 287483.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.2687 -0.6669  0.0180  0.6754  3.4649

Random effects:
 Groups   Name        Variance Std.Dev.
 SCHOOLID (Intercept)  703.2   26.52
 Residual             6084.5   78.00
Number of obs: 24775, groups:  SCHOOLID, 1040

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   504.2892     1.0653   979.1664 473.365  < 2e-16 ***
BELONG          4.7224     0.4583 24310.7359  10.304  < 2e-16 ***
DISCLISCI       2.8430     0.5983 23755.9526   4.752 2.03e-06 ***
TEACHSUP        4.2959     0.5712 23744.9846   7.521 5.64e-14 ***
gmDISCLISCI    20.7541     2.7842  1206.2872   7.454 1.72e-13 ***
gmTEACHSUP     -7.9306     3.2917  1173.6108  -2.409   0.0161 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) BELONG DISCLI TEACHS gDISCL
BELONG      -0.052
DISCLISCI    0.005 -0.099
TEACHSUP     0.006 -0.123 -0.209
gmDISCLISCI  0.113  0.001 -0.214  0.047
gmTEACHSUP  -0.370  0.012  0.037 -0.172 -0.399
```

c) Assign the estimates (i.e., their values) in the output to the elements of this model.
Note: For random effects, report one decimal; for fixed effects, report 4 decimals.

Elements of the model	Estimate in the R output
c_{00}	504.2892
σ_r^2	6084.5
$\sigma_{u_0}^2$	703.2
c_{01}	20.7541
c_{02}	-7.9306
b_1	2.8430
b_2	4.2959
b_3	4.7224

SCORING:

- 0.5 credits per correct assignment
- **Subtotal: 4 credits**

d) Report the **two contextual effects** of teacher support and disciplinary climate.
Note: Report 4 decimals for each of the components below.

Contextual effect	Estimate	SE	p-value
Teacher support	-7.9306	3.2917	0.0161
Disciplinary climate	20.7541	2.7842	0.0000

SCORING:

- 1 credit per row
- **Subtotal: 2 credits**

e) The researchers decided to **group mean-center the L1 predictors** TEACHSUP and DISCLISCI in the model.

Does this centering change the computation of the contextual effects? If yes, how? If no, why not?

Computation of the contextual effect
Yes, it changes the way the contextual effect is computed. If c_{01} denotes the L2 effect of gmDISCLISCI and b_1 the L1 effect of the group mean-centered variable DISCLISCI, then the contextual effect is computed as $c_{01} - b_1$ (similar for TEACHSUP).

SCORING:

- 1 credit for acknowledging that there is a change
- 1 credit for describing the change
- **Subtotal: 2 credits**

f) Formulate a research question that may underlie the model.

Research question
For instance: To what extent do school disciplinary climate and teacher support explain variation in individual students' mathematics achievement after controlling for their individual perceptions of disciplinary climate and teacher support and sense of belonging?

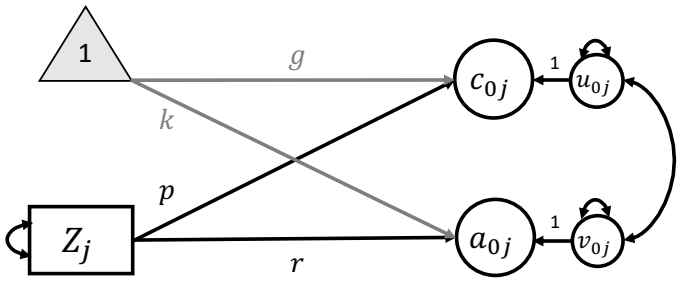
SCORING:

- 1 credit for specifying that the L2 variables relate to the L1 variable
- 1 credit for acknowledging the L1 control variables
- **Subtotal: 2 credits**

A3. Multilevel Path Models (19 credits)

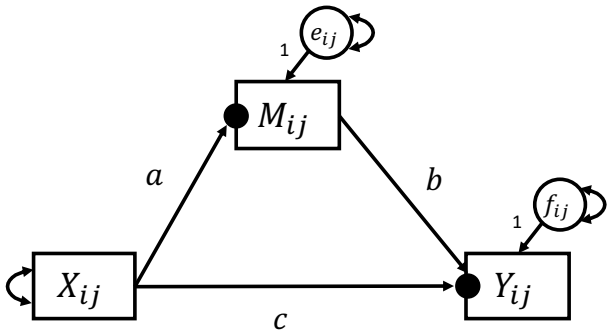
The following figure shows a multilevel path model (**Model 1**).

Model 1



L2
Between level

L1
Within level



This model contains three variables at the within level (X, M, and Y) which are assumed to follow a mediation model. The random intercept of Y is labelled c_{0j}, and a_{0j} for the variable M. The variable Z is a level-2 predictor variable.

Note: The index i represents the within level (L1), the index j the between level (L2).

a) Write out the model equations for the within level and the between level.

Within-level equations:

Dependent variables	Model equation
$M_{ij} =$	$a_{0j} + aX_{ij} + e_{ij}$
$Y_{ij} =$	$c_{0j} + cX_{ij} + bM_{ij} + f_{ij}$

Between-level equations:

Dependent variables	Model equation
$c_{0j} =$	$g + pZ_j + u_{0j}$

$a_{0j} =$	$k + rZ_j + v_{0j}$

SCORING:

- 1 credit per correct model equation
- **Subtotal: 4 credits**

b) Which of the following statements about the model are true?

Statement	True	False
(1) The model contains two random slopes a_0 and c_0, and one random intercept c_1.	☐	X
(2) The model contains a cross-level interaction effect.	☐	X
(3) The model controls for within-level variation in the variable Z.	☐	X
(4) The L2 random effects are assumed to be independent.	☐	X

NOTES:

- Statement (1): FALSE. a_0 and c_0 are random intercepts (of M and Y), and c_1 is the fixed slope (of the relation between X and Y).
- Statement (2): FALSE. The model does not contain any random slope which is further explained by a between-level variable.
- Statement (3): FALSE. The variable Z is only measured at the between level (L2). Hence, it does not have a within-level variation.
- Statement (4): FALSE. The figure indicates that the L2 random effects can covary.

SCORING:

- 1 credit per correct response
- **Subtotal: 4 credits**

c) Which type of multilevel mediation model does Model 1 represent?

Model type	Response
1-2-2	☐
2-2-1	☐
1-2-1	☐
2-1-1	☐
1-1-1	X
2-1-2	☐

NOTE:
All three variables (i.e., the predictor X, the mediator M, and the outcome variable Y) were measured at the within-level (i.e., at L1).

SCORING:

- 1 credit per correct response
- **Subtotal: 1 credit**

d) Provide the lavaan syntax that specifies Model 1 and estimates the indirect and total effect at L1.

Note: Use the variable names and the names of the regression coefficients from the figure.

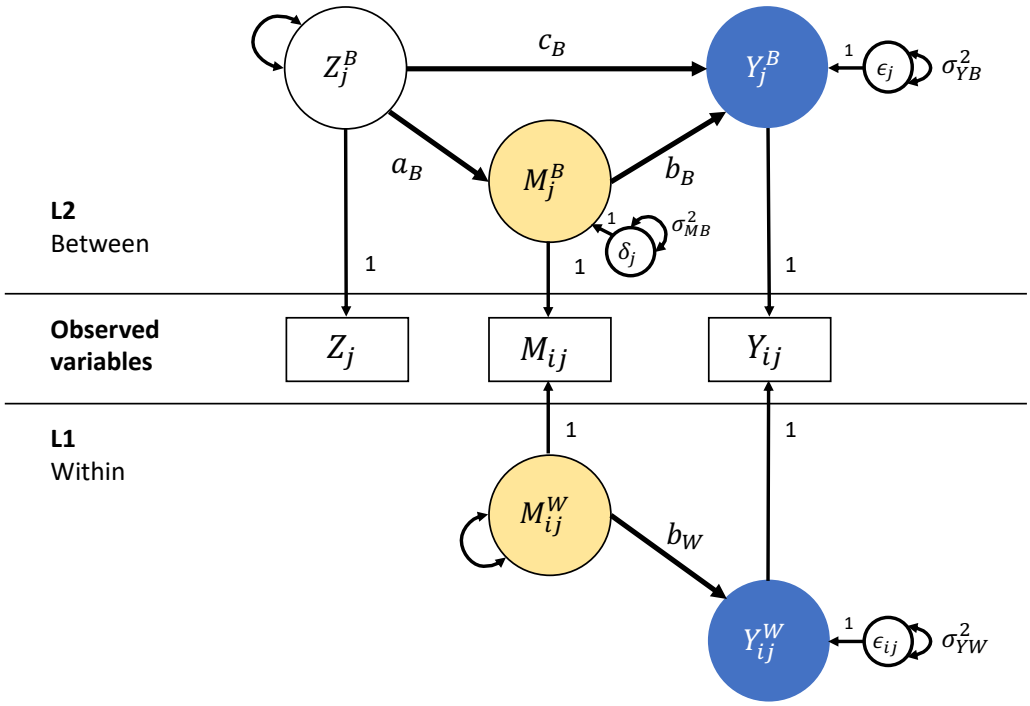
Lavaan syntax
Model1 <- ' level: 1 Y ~ c1*X + b*M M ~ a*X level: 2 Y ~ p*Z M ~ r*Z Y ~~ M # Indirect effect at L1 ind := a*b # Total effect at L1 tot := a*b + c1 '

SCORING:

- 1 credit per correct syntax line
- **Subtotal: 7 credits**

e) Multilevel path models can also be represented as multilevel structural equation models (MSEM). Some researcher has decided to drop the variable X in Model 2 and assume a mediation at L2 rather than L1. The new model (named **Model 2**) is represented in the MSEM framework below.

Model 2



Which type of multilevel mediation model does Model 2 represent?

Model type	Response
1-2-2	☐
2-2-1	☐
1-2-1	☐
2-1-1	X
1-1-1	☐
2-1-2	☐

How is the **indirect L2 effect** computed in this model?

Computation of the indirect L2 effect
$ind_{L2} = a_B b_B$

How is the contextual effect of the L2 mediator M on the L1 outcome variable Y computed in this model?

Computation of the contextual effect
$cont = b_B - b_W$

SCORING:

- 1 credit per correct response

- **Subtotal: 3 credits**

D1. Implementing multilevel models in R (14 credits)

The R packages lme4 and lavaan provide several options to specify and estimate multilevel models. In the following, the PISA 2006 data (pisa06) are used to illustrate these options.

Variables in the data set:

- math: Individual students' mathematics achievement score
- schoolid: School identifying variable (ID)
- ses: Individual students' socioeconomic status
- read: Individual students' reading achievement score
- schoolses: School-average socioeconomic status

a) Explain the four snippets of syntax by naming the **types of the models** they specify and determine whether the **intercepts and slopes are fixed or random**.

Syntax snippets	
(1)	lmer(math ~ 1 + (1|schoolid), data = pisa06)
(2)	lmer(math ~ 1 + ses + (1|schoolid), data = pisa06)
(3)	lmer(math ~ 1 + ses + read + (1 + ses|schoolid), data = pisa06)

(4)	Model2 <- '
	  level: 1
	  math ~ ses
	
	  level: 2
	  math ~ schoolses
	'

Syntax		Explanation
(1)	lmer(math ~ 1 + (1|schoolid), data = pisa06)	▪ Two-level null model ▪ Random intercepts of math at L2
(2)	lmer(math ~ 1 + ses + (1|schoolid), data = pisa06)	▪ Two-level random-intercept model ▪ Random intercepts of math at L2

(3)	lmer(math ~ 1 + ses + read + (1 + ses|schoolid), data = pisa06)	▪ Two-level random-intercept-and-random-slope model ▪ Random intercepts of math at L2 ▪ Random slopes of the relation between math and ses

(4)	Model2 <- '
	  level: 1
	  math ~ ses
	
	  level: 2
	  math ~ schoolses
	'
		▪ Contextual model
		▪ Contextual effect of school-average SES on individual mathematics achievement
		▪ Random intercepts of math at L2

SCORING:

- 1 credit per correct type of model
- 1 credit per correct specification of the L2 random effects
- **Subtotal: 8 credits**

b) Provide the L1 and L2 model equations for syntax snippet (3).

Model specification
L1 (students): $$math_{ij} = \beta_{0j} + \beta_{1j}ses_{ij} + \beta_{2j}read_{ij} + r_{ij}$$ L2 (schools): $$\beta_{0j} = \gamma_{00} + u_{0j}$$ $$\beta_{1j} = \gamma_{10} + u_{1j}$$ $$\beta_{2j} = \gamma_{20}$$

SCORING:

- 1 credit for the correct specification of Y as the outcome and the predictors
- 1 credit for the correct indices at L1
- 1 credit for the correct specification of the L2 equation
- **Subtotal: 3 credits**

c) Which of the snippets of syntax can be used to address the following three research questions?

Research Question	Syntax snippet			
	(1)	(2)	(3)	(4)
(1) To what extent does the relation between individual students' mathematics achievement and socioeconomic status vary between schools after controlling for individual reading skills?	☐	☐	X	☐
(2) How much variance does socioeconomic status explain in mathematics achievement at the school level after controlling for student-level variation in these variables?	☐	☐	☐	X
(3) How large is the contextual effect of school SES on individual achievement in math?	☐	☐	☐	X

SCORING:

- 1 credit per correct response
- **Subtotal: 3 credits**

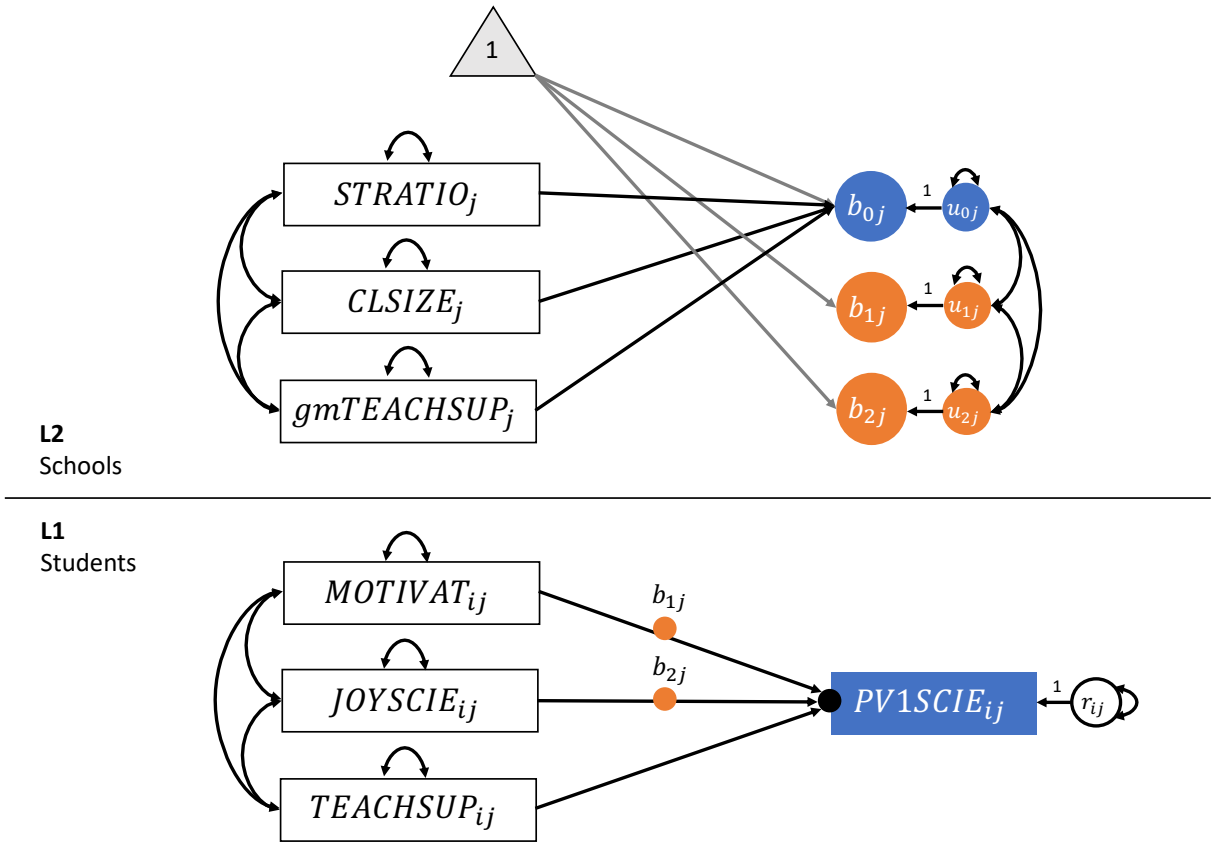
D2. Enjoyment and Motivation in Science (10 credits)

The Programme for International Student Achievement (PISA) 2015 assessed science achievement for a large sample of 15-year-old students in the Nordic countries. The data (pisa15) contain the following variables:

- SCHOOLID: Identification variable of the school
- PV1SCIE: Individual students' science achievement (test score)
- MOTIVAT: Individual students' science motivation
- JOYSCIE: Individual students' enjoyment of science
- TEACHSUP: Individual students' perceptions of the teacher support
- STRATIO: Student-teacher ratio in the school
- CLSIZE: Average class size in the school
- gmTEACHSUP: Group mean of students' perceptions of the teacher support

All variables were positively coded so that higher scores indicate better achievement, more motivation and enjoyment, more teacher support, a larger student-teacher ratio, and larger classes.

A researcher wanted to test the following two-level model:



a) How can the researcher implement this model in lme4 (named Model1)? Formulate the model specification syntax. You may name the cluster variable "SCHOOLID".

Model specification syntax
Model1 <- lmer(PV1SCIE ~ 1 + JOYSCIE + MOTIVAT + TEACHSUP + gmTEACHSUP + STRATIO + CLSIZE + (1 + MOTIVAT + JOYSCIE | SCHOOLID) , data = pisa15)

SCORING:

- 1 credit for correctly specifying the outcome variable
- 1 credit for correctly specifying the L1 predictors
- 1 credit for correctly specifying the L2 predictors
- 1 credit for correctly specifying the random intercept (1)
- 1 credit for correctly specifying the random slopes
- **Subtotal: 5 credits**

b) How many random effects does this model have? How many elements describe the structure of the L2 random effects?

Number of random effects:	4
Structure of the L2 random effects:	3 or 6

NOTES:

- Random effects (variances L1 and L2 residuals): σ_r^2, $\sigma_{u_0}^2$, $\sigma_{u_1}^2$, $\sigma_{u_2}^2$
- Structure of the L2 random effects (covariances among L2 residuals): $\sigma_{u_0u_1}$, $\sigma_{u_0u_2}$, $\sigma_{u_1u_2}$
- *If students named 6 elements in the structure of the random effects, they gained the credit, too. The term may have been unclear, and "structure" could be interpreted as comprising all elements in the L2 covariance matrix of the random effects (including the variances).*

SCORING:

- 1 credit per correct number
- **Subtotal: 2 credits**

c) A researcher estimated this model and obtained the following fixed effects parameters:

```
Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)  5.050e+02  8.646e+00  7.759e+02  58.411  < 2e-16 ***
JOYSCIE      2.377e+01  6.335e-01  5.321e+02  37.515  < 2e-16 ***
MOTIVAT      1.217e+01  6.879e-01  1.847e+03  17.684  < 2e-16 ***
TEACHSUP     1.941e+00  7.402e-01  1.597e+04   2.622  0.00874 **
STRATIO      1.243e-01  5.333e-01  7.600e+02   0.233  0.81571
CLSIZE      -8.786e-02  3.446e-01  7.138e+02  -0.255  0.79883
gmTEACHSUP   7.873e+00  4.822e+00  8.087e+02   1.633  0.10293
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Which of the following conclusions can be drawn (i.e., are true)?

Conclusion	True	False
(1) Individual students' perceptions of teacher support (TEACHSUP) do not explain variation in their science achievement.	☐	X
(2) Individual students' enjoyment of science and motivation are positively and significantly related to science achievement after controlling for TEACHSUP, STRATIO, CLSIZE, and the school-average TEACHSUP (gmTEACHSUP).	X	☐
(3) There is evidence supporting the contextual effects of STRATIO, CLSIZE, and school-average TEACHSUP (gmTEACHSUP).	☐	X

NOTES:

- (1): FALSE, because there is evidence for the significant relation between these two variables at L1 (B = 1.941, SE = 0.740, p = 0.009).
- (2): TRUE.
- (3): FALSE, because none of these L2 variables are significantly related to the L1 science achievement.

SCORING:

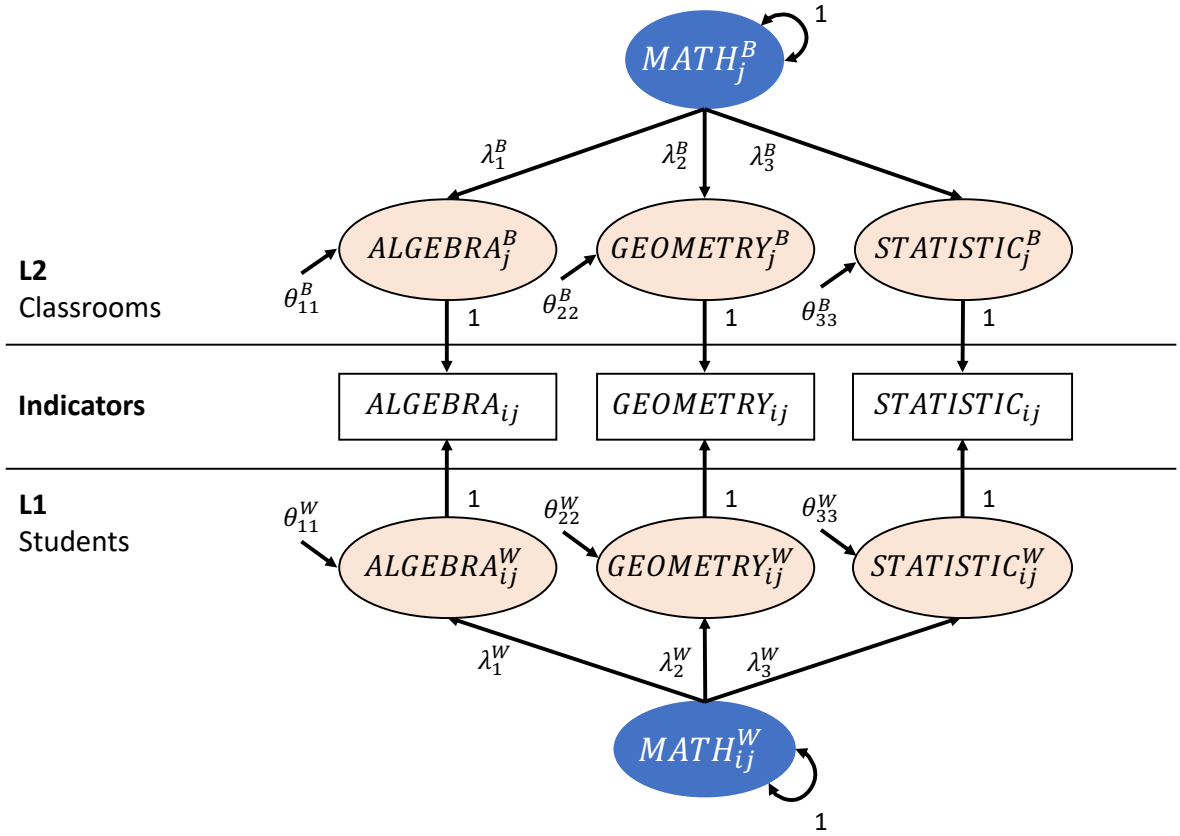
- 1 credit per correct response
- **Subtotal: 3 credits**

D3. Multilevel Confirmatory Factor Analysis (6 credits)

Multilevel confirmatory factor analysis enables researchers to evaluate scales at different levels of analysis—this includes the level-specific reliability.

A researcher has collected data from students in classrooms ($ClassID$) comprising mathematics test scores in algebra ($ALGEBRA$), geometry ($GEOMETRY$), and statistics ($STATISTIC$). He or she assumes that the latent (unobserved) variable representing mathematics skills ($MATH$) is indicated by these three manifest (observed) scores.

Given that student data are nested in classroom data, the researcher specifies a multilevel confirmatory factor analysis model as follows:



Estimating this model in the R package lavaan, the researcher obtained the following output:

```
## Level 1 [within]:
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   MATH =~
##     ALGEBRA           1.000                               3.134    0.779
##     GEOMETRY          1.006    0.077   13.078    0.000    3.152    0.806
##     STATISTIC         0.975    0.076   12.907    0.000    3.056    0.773
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ALGEBRA           0.000                               0.000    0.000
##    .GEOMETRY          0.000                               0.000    0.000
##    .STATISTIC         0.000                               0.000    0.000
##     MATH              0.000                               0.000    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ALGEBRA           6.369    0.750    8.489    0.000    6.369    0.393
##    .GEOMETRY          5.352    0.710    7.535    0.000    5.352    0.350
##    .STATISTIC         6.302    0.729    8.645    0.000    6.302    0.403
##     MATH              9.822    1.279    7.680    0.000    1.000    1.000
##
## R-Square:
##                    Estimate
##     ALGEBRA           0.607
##     GEOMETRY          0.650
##     STATISTIC         0.597

## Level 2 [ClassID]:
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   MATH =~
##     ALGEBRA           1.000                               3.121    0.958
##     GEOMETRY          0.944    0.109    8.673    0.000    2.946    0.909
##     STATISTIC         0.863    0.104    8.268    0.000    2.692    0.897
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ALGEBRA          29.884    0.469   63.783    0.000   29.884    9.175
##    .GEOMETRY         29.882    0.464   64.413    0.000   29.882    9.225
##    .STATISTIC        29.736    0.437   68.005    0.000   29.736    9.911
##     MATH              0.000                               0.000    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ALGEBRA           0.870    0.709    1.226    0.220    0.870    0.082
##    .GEOMETRY          1.814    0.752    2.411    0.016    1.814    0.173
##    .STATISTIC         1.755    0.701    2.506    0.012    1.755    0.195
##     MATH              9.739    2.496    3.901    0.000    1.000    1.000
##
## R-Square:
##                    Estimate
##     ALGEBRA           0.918
##     GEOMETRY          0.827
##     STATISTIC         0.805
```

Note. Estimate: Unstandardized estimate of the parameter with standard error (Std.Err); Std.lv: Standardized estimate where only the latent variable (lv) was standardized; Std.all:

Fully standardized estimate where both the latent variable and the indicators were standardized.

a) Given that a measurement model for $MATH$ is specified at both the student and the classroom level, two reliabilities are available: the student-level reliability (ω_{L1}) and the classroom-level reliability (ω_{L2}). These reliabilities are computed based on the fully standardized estimates as follows:

$$\omega_{\mathrm{L1}} = \frac{(\lambda_1^W + \lambda_2^W + \lambda_3^W)^2}{(\lambda_1^W + \lambda_2^W + \lambda_3^W)^2 + \theta_{11}^W + \theta_{22}^W + \theta_{33}^W}$$

$$\omega_{\mathrm{L2}} = \frac{(\lambda_1^B + \lambda_2^B + \lambda_3^B)^2}{(\lambda_1^B + \lambda_2^B + \lambda_3^B)^2 + \theta_{11}^B + \theta_{22}^B + \theta_{33}^B}$$

Given the output and the equations, how do you compute these two reliabilities? Please provide the calculations. There is no need to compute the final values.

Reliability	Computation
$\omega_{\mathrm{L1}} =$	$\frac{(0.779 + 0.806 + 0.773)^2}{(0.779 + 0.806 + 0.773)^2 + 0.393 + 0.350 + 0.403}$
$\omega_{\mathrm{L2}} =$	$\frac{(0.958 + 0.909 + 0.897)^2}{(0.958 + 0.909 + 0.897)^2 + 0.082 + 0.173 + 0.195}$

SCORING:

- 1 credit for choosing the fully standardized estimates in the two equations
- 1 credit per correct equation
- **Subtotal: 3 credits**

The lavaan syntax to specify this model (named mcfa1) is shown below.

Lavaan syntax

```
mcfa1 <- '
        level: 1
        MATHw =~ ALGEBRA + GEOMETRY + STATISTIC
        level: 2
        MATHb =~ ALGEBRA + GEOMETRY + STATISTIC
        '
```

To evaluate the fit of such a model, the L1 or L2 parts can be modified so that they represent null, independence, or saturated models. In the course, we have also encountered these types of models in the context of multilevel path models.

b) Modify this syntax to estimate the model with a **L2 saturated model** (named mcfa2).

Lavaan syntax

```
mcfa2 <- '
                    level: 1
                    MATHw =~ ALGEBRA + GEOMETRY + STATISTIC
                    level: 2
                    ALGEBRA ~~ GEOMETRY + STATISTIC
                    GEOMETRY ~~ STATISTIC
                    '
```

Alternative specification:

Lavaan syntax

```
mcfa2 <- '
                    level: 1
                    MATHw =~ ALGEBRA + GEOMETRY + STATISTIC
                    level: 2
                    ALGEBRA ~~ ALGEBRA  + GEOMETRY + STATISTIC
                    GEOMETRY ~~ GEOMETRY + STATISTIC
                    STATISTIC ~~ STATISTIC
                    '
```

Scoring:

- 1 credit for the correct L1 model
- 2 credits for the correct L2 model (covariances ~~ and variables)
- **Subtotal: 3 credits**