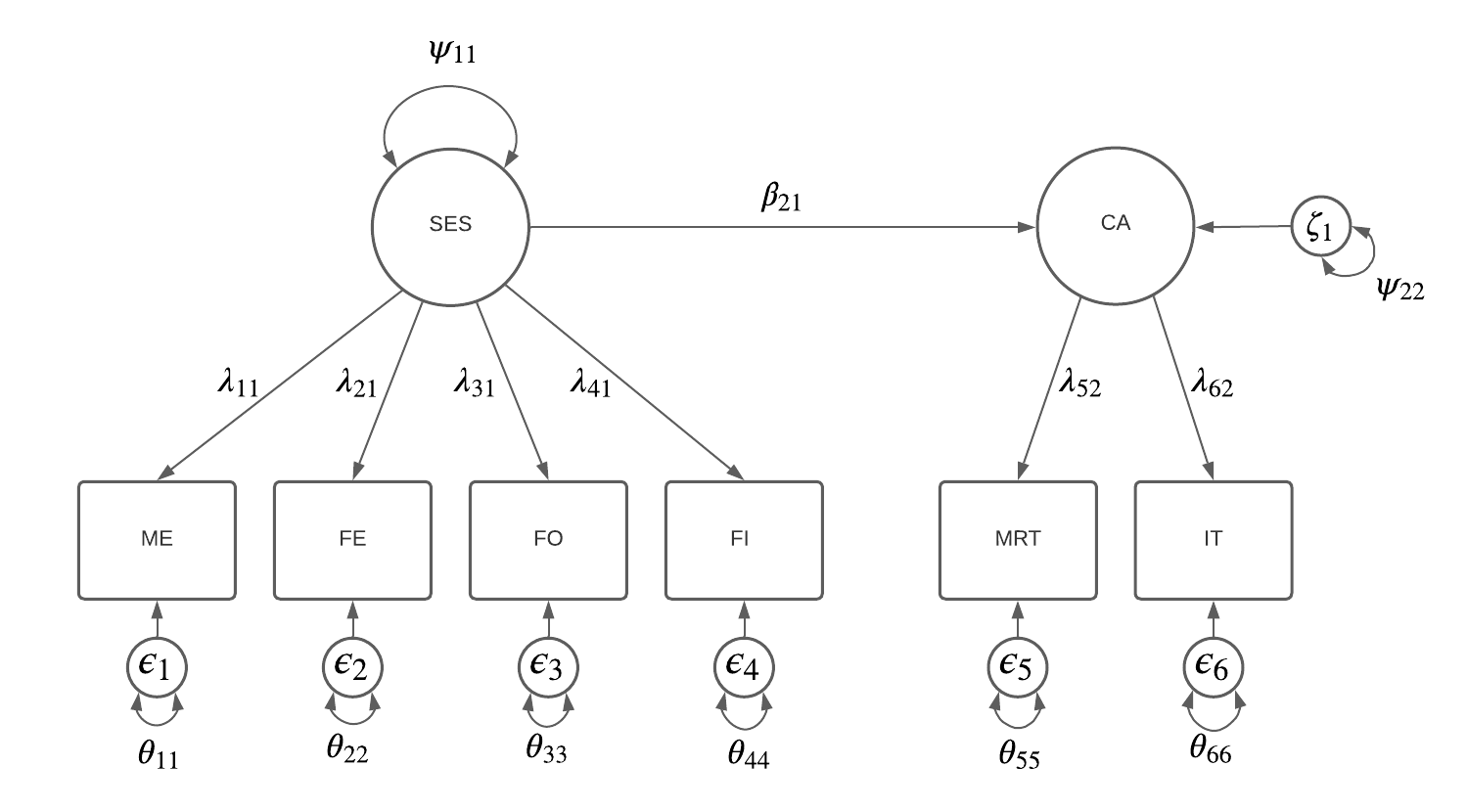
**Hand-in exercise 1**

**Please print and handwrite the answers**

Head Start is a program of the United States Department of Health and Human Services that provides comprehensive education, health, nutrition, and parent involvement services to low-income children and their families. Magidson (1977) obtained data from 148 children who participated in the Head Start Summer Program, and 155 children who form a control group. Here we focus on the low-income children only.

The children were measured on six variables: Mother’s education; Father’s education; Father’s occupation; Family Income; Metropolitan Readiness Test; and the Illinois Test for Psycholinguistic Abilities. We assume the first four variables are indicators of the latent variable (i.e., factor) Socio-economic Status, while the latter two variables are indicators of the latent variable Cognitive Ability. Furthermore, we will try to predict Cognitive Ability from Socioeconomic Status.

a. Draw the path diagram for this model.



b. How many sample statistics are there (which ones)?

*The number of sample statistics is calculated by adding up the elements in the covariance matrix, and the elements in the mean structure. Because we have six items, the mean structure contains sample statistics. The number of elements in the covariance matrix is given by:*

*The total number of sample statistics adds up to .*

c. How many parameters are there in the model (which ones)? So how many df are there?

*When we would use unit loading identification to identify our model, we would freely estimate the following parameters:*

*Factor loadings:*

*Variances of the residuals:*

*Means:*

*Regression of CA on SES:*

*Variances of SES:*

*Variance of the residuals for CA:*

*In total: parameters*

*The degrees of freedom are calculated by subtracting the number of freely estimated parameters from the number of sample statistics. Therefore, we have degrees of freedom.*

d. The data are given in the file **HeadStartLow.dat**. Note that is does not contain the raw data, but only the summary statistics, that is, the means, standard deviations and (lower triangular) correlation matrix. You need to tell Mplus how to read this data file. Use:

DATA: TYPE IS MEANS STDEVIATIONS CORRELATION;

FILE IS HeadStartLow.dat;

NOBSERVATIONS ARE 147;

VARIABLE: NAMES = MoEdu FaEdu FaOcc FamIn Ready PsAb;

Can the covariance matrix be obtained from these summary statistics? How?

*The covariance matrix can be derived from the correlations and standard deviations. Correlations are calculated by dividing the covariance between x and y by the product of the standard deviations of x and y. Therefore, we can derive the covariance matrix with the following expression:*

*Applying this formula to the entire correlation matrix yields the following covariance matrix:*

*This covariance matrix can also be obtained by specifying the model and including ‘sampstat’ in the output.*

e. Specify the model. To regress Cognitive Ability on Socioeconomic Status, use the statement ON (see chapter 17 of the Mplus User’s Guide). How did you specify the model?

TITLE: Specification of simple latent regression model,

CA is predicted by SES at the level of the latent variables.

Factor SES is measured by ME, FE, FO and FI

Factor CA is measured by MRT and the IT

DATA: TYPE IS MEANS STDEVIATIONS CORRELATION;

FILE IS HeadStartLow.dat;

NOBSERVATIONS ARE 147;

VARIABLE: NAMES = MoEdu FaEdu FaOcc FamIn Ready PsAb;

MODEL: SES by MoEdu FaEdu FaOcc FamIn;

CA by Ready PsAb;

CA on SES;

OUTPUT: sampstat; TECH1;

f. Ask for the TECH1 output to make sure you specified the model correctly. Where in the TECH1 output is the parameter by which Cognitive Ability is regressed on Socioeconomic Status?

*The parameter by which cognitive ability is regressed on socio-economic status is . In the TECH1 output, it can be found in the matrix called BETA. The parameter is underlined below. As can be observed, this parameter is the seventeenth to be freely estimated.*

BETA

SES CA

\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_

SES 0 0

CA 17 0

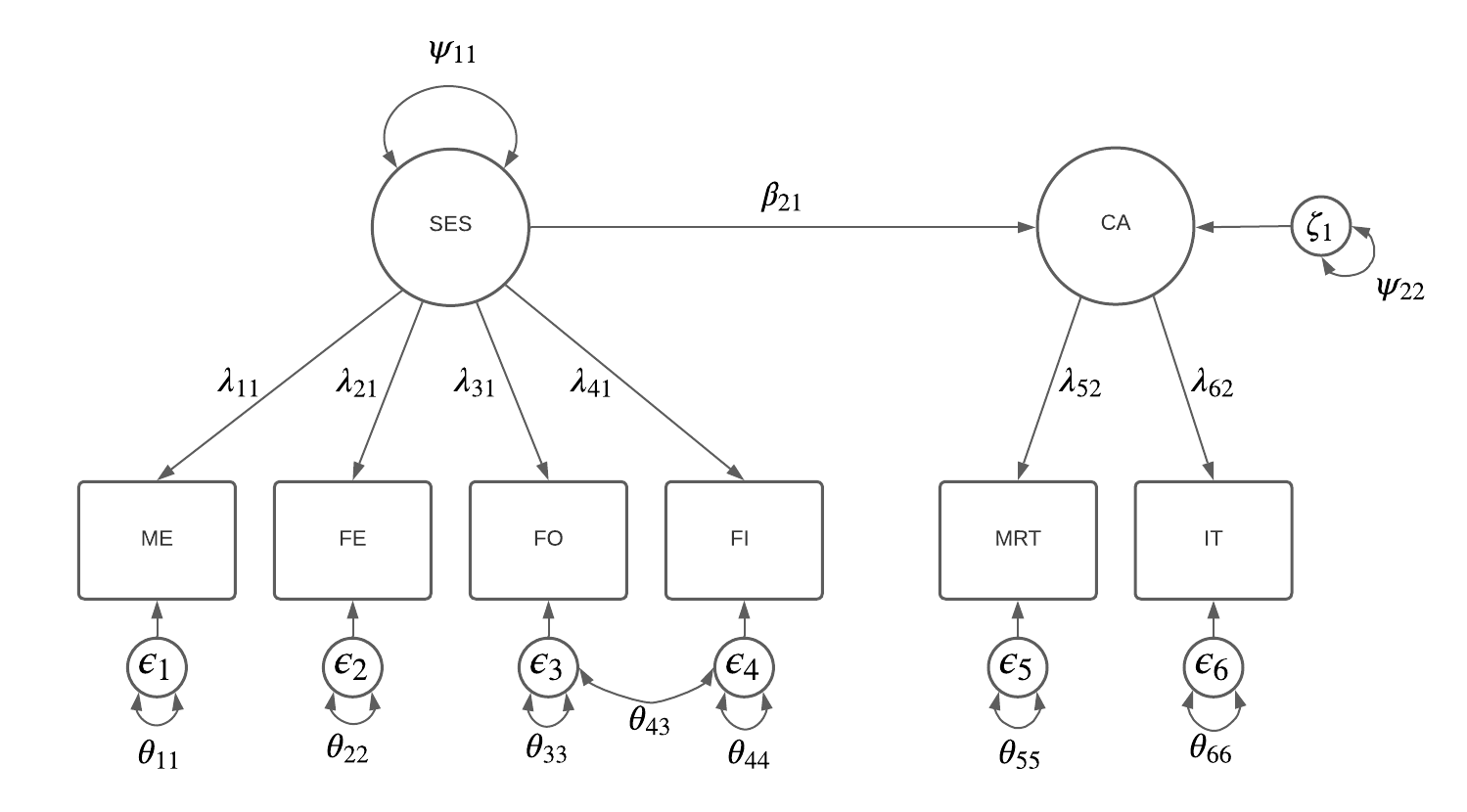
g. Discuss the model fit.

*There are multiple measures by which the model fit can be assessed. Below, we give the value for each fit measure accompanied by an interpretation.*

* *Goodness of fit test: The goodness of fit test is significant (. This indicates that our model is rejected in favor of the saturated model. According to the goodness of fit test, our model does not fit well.*
* *RMSEA: The RMSEA is . The RMSEA should be smaller than . Since this is not the case, our model does not fit well according to this measure*
* *CFI and TLI: The CFI is and the TLI is . Both measures should be over 0.90 for a moderate model fit, and over 0.95 for a good model fit. According to the CFI, the fit is moderate, according to the TLI, the fit is bad*
* *SRMR: The SRMR is . In order to have a good model fit, SRMR should be under 0.08. Therefore, we conclude that there is a good model fit according to this measure.*

*All in all, we would conclude that the model does not fit well to the data, since the majority of the fit measures indicates a bad fit.*

h. Run the model again including MOD(4) on the OUTPUT line. According to the modification indices, which parameter would lead to the largest improvement? Indicate in the path diagram where this parameter would be. Does this model make sense from a substantive perspective? (Remember: these data were obtained in the 70s.)



i. Run the modified model, discuss the model fit, and compare it to the previous model. What can you conclude? Is this what you expected?

j. Can Cognitive Ability be predicted from Socioeconomic Status? By how much (TIP: Use standardized results option)?

k. Write down the expressions for each of the observed variables in terms of SES.

l. Write down the covariance between Father’s occupation and Family income.

m. Do the same for Mother’s education and Psycholinguistic Abilities.