Introduction

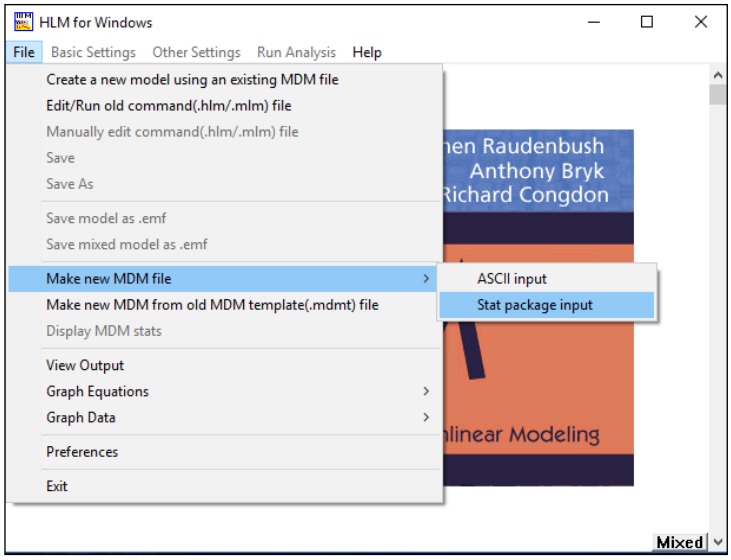
This document shows how you can replicate the popularity data multilevel models from Multilevel analysis: Techniques and applications (Hox, J. J., Moerbeek, M., & van de Schoot, R, 2018), Chapter 2. In this manual the software package HLM7 for Windows was used. Results should be very similar to results obtained with other software packages, however due to convergence and rounding issues, you might notice minor differences. This tutorial is based on earlier versions of similar manuals and made by Laurent Smeets.

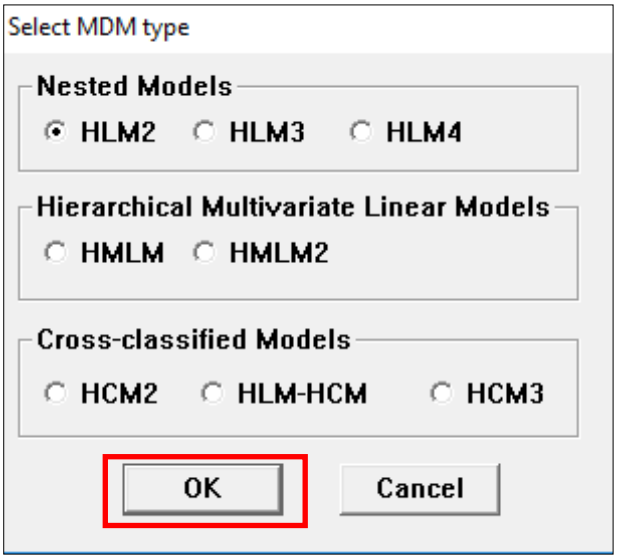
STEP 1: Downloading the data.

To download the popularity data go to https://multilevel-analysis.sites.uu.nl/datasets/ and follow the links to https://github.com/MultiLevelAnalysis/Datasets-third-edition-Multilevel-book/blob/master/chapter%202/popularity/SPSS/popular2.sav. We will use the .sav file in HLM, which can be found in the SPSS folder.

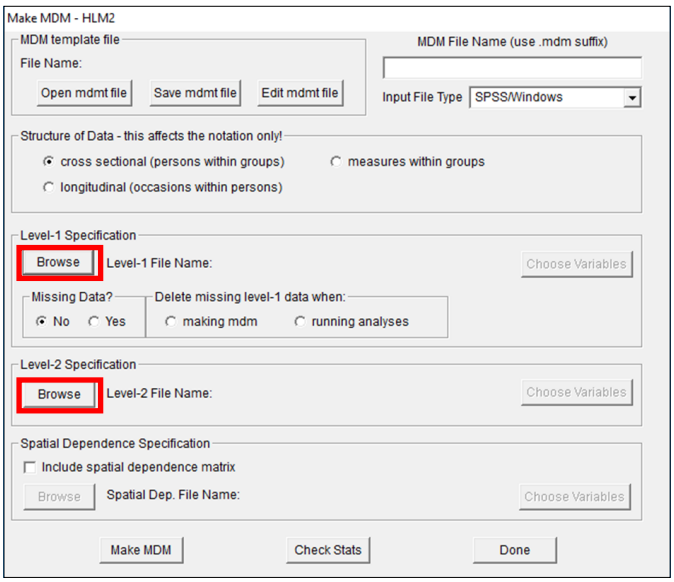
STEP 2: Loading the data.

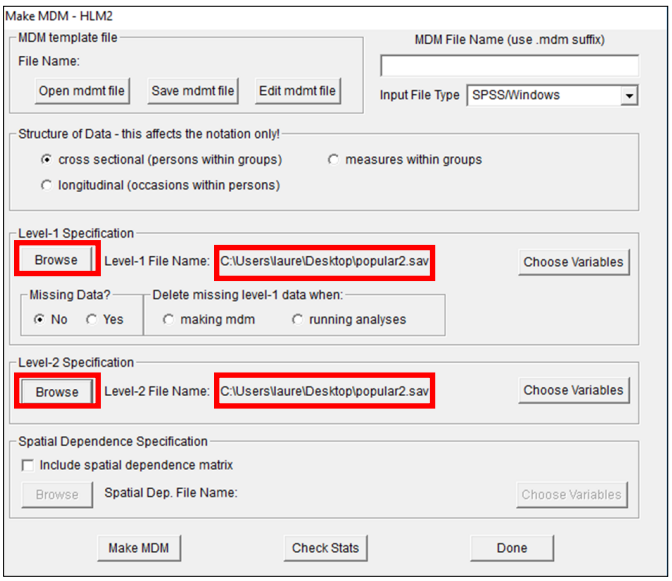
1) To load the SPSS datafile go to **File > Make new MDM file > Stat package input**. In the new window that appears, select **HLM2** and press **OK**.





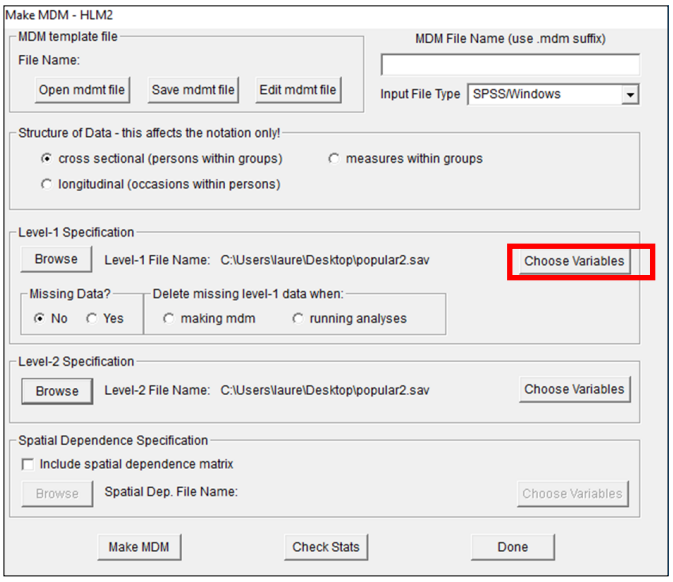
Both our level 1 and level 2 variables are in one and the same file (they can also be in two separate files as long as there is a common identifying variable for the level 2 units). Therefore, in the new window that appears, select the popular2.sav data-file for both levels (In this example level 1 and 2 by pressing **Browse**). Now simply look for the file and select it.

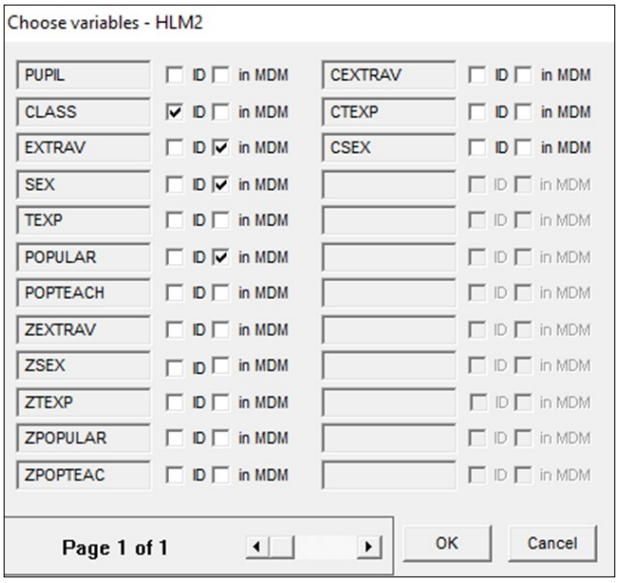




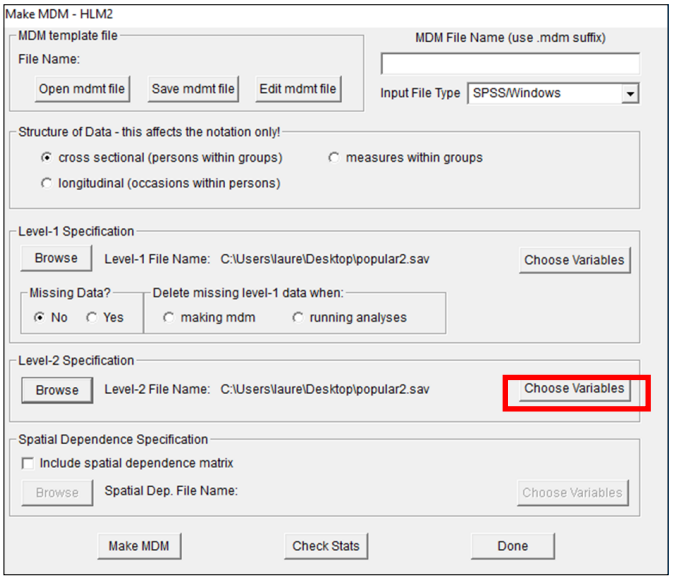
Next, we specify the 'grouping' variable (or identification variable), and which variables belong to which level. We start with **level one**. Press **Choose Variables** to open the variables window. The 'grouping' variable (ID), in a two level analysis is the identification variable at the second level (in a three level model there are two ID variables: one on the second level and one on the third level). There can only be one grouping variable per level and the first level (in this example the pupil level) does not need an ID variable. The ID variable is the only variable that we need to select in both level 1 and level 2. Since our second level is the class level, the CLASS variable is our ID variable. All the other variables are selected only once together with the level at which they were measured. In the window that appears we select the grouping variable and the level 1 variables. The grouping variable is **CLASS**. Therefore we select the **ID checkbox** for this variable. The level 1 variables are **EXTRAV**, **SEX** and **POPULAR**. Select the **in MDM** checkbox for these variables. When you are done, press **OK**.

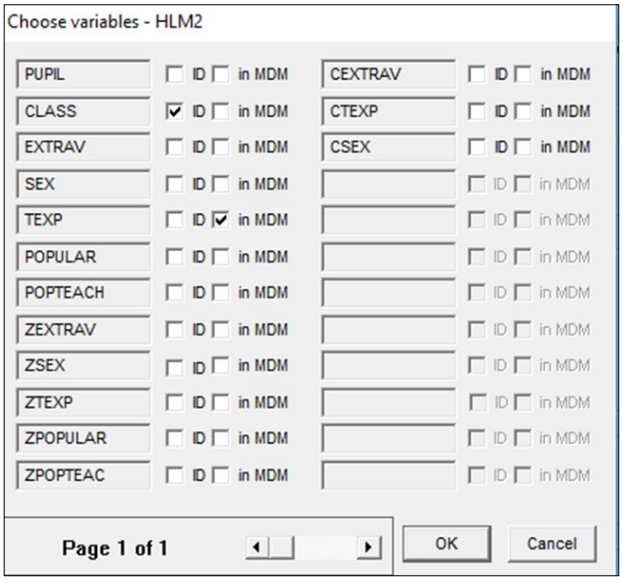
The dataset also includes Z-transformed variables, but these will be ignored in this example.



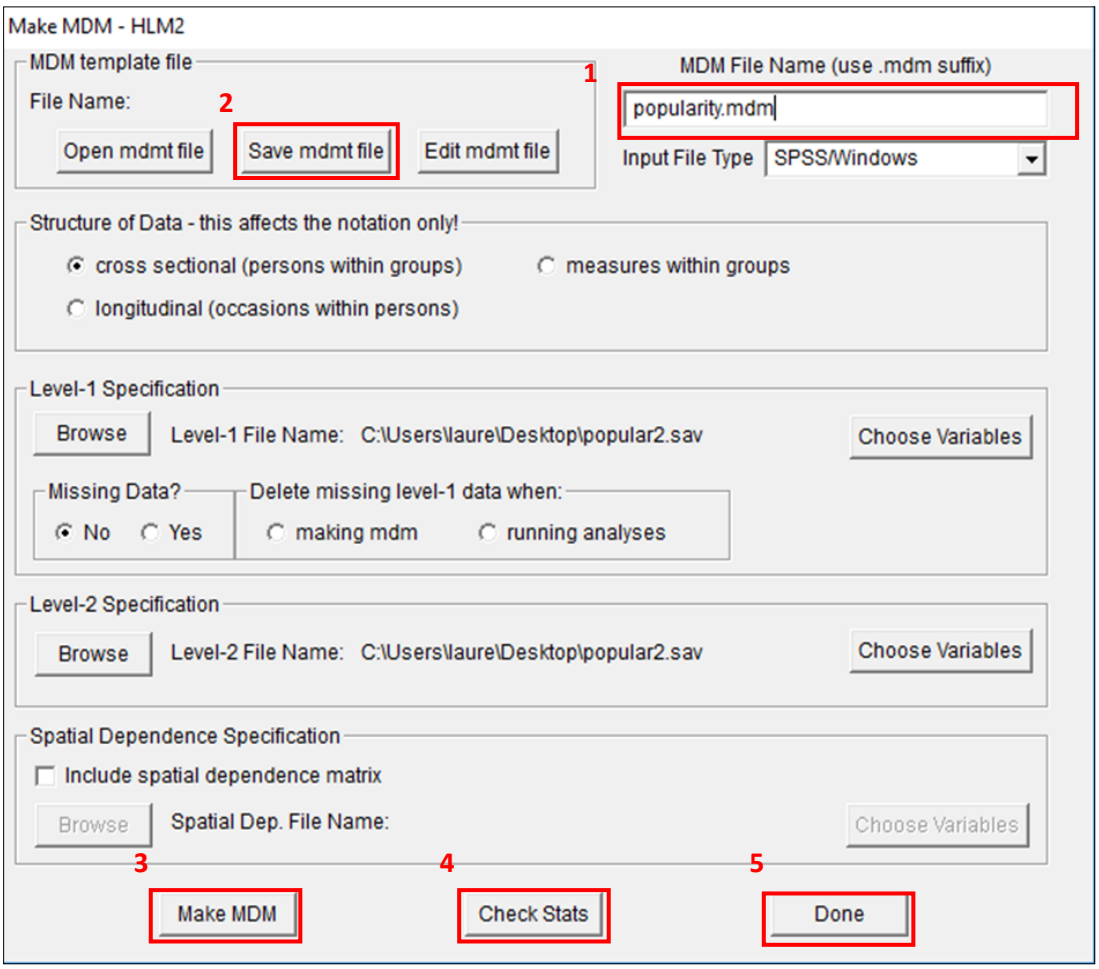


In a similar way, we select the variables for Level 2. Press the lower **Choose Variables** button to specify level 2 variables.

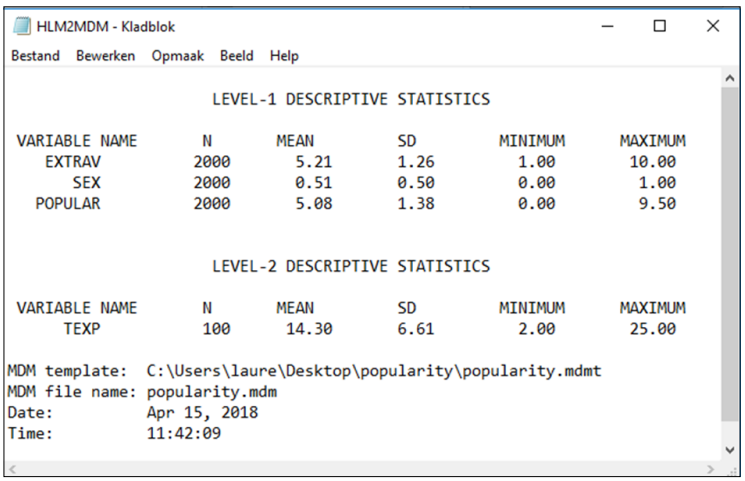


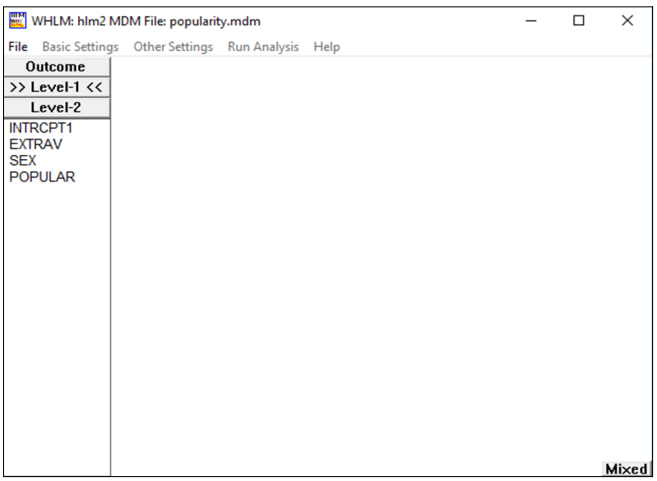


Once all variables have been specified, it is time to save the file. First specify a **filename** in the upper right box, and then press **Save mdmt file**. HLM will create multiple files (often with the same name) while doing your analysis, it is thus advised to create a new folder for each project you are working on, to prevent cluttering and confusion.



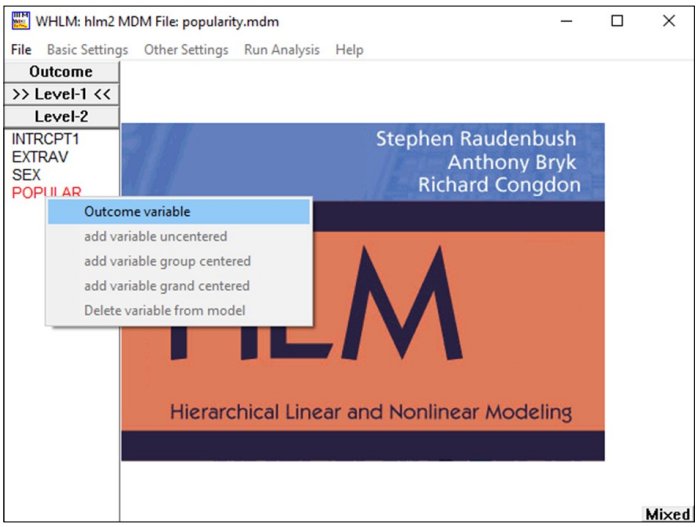
After the file has been saved successfully, you can generate the actual multilevel file. First press **Make MDM**. This will show a terminal window for a few seconds. Then Press **Check Stats**. This will open a text file with some descriptive statistics from your variables. Verify that these are correct, and then press **Done**.

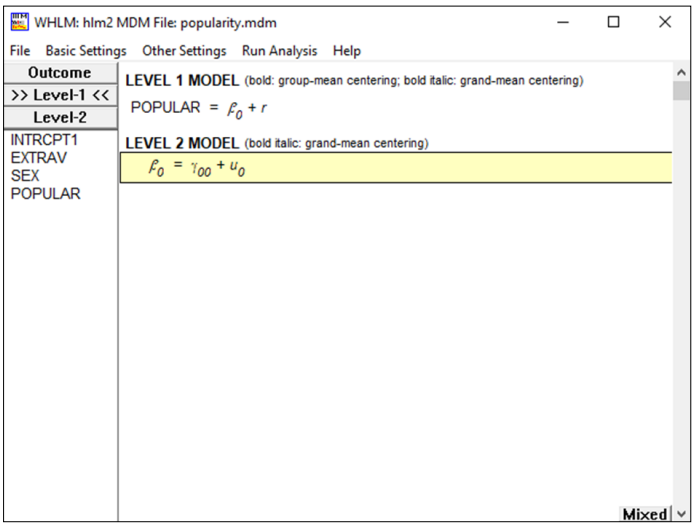




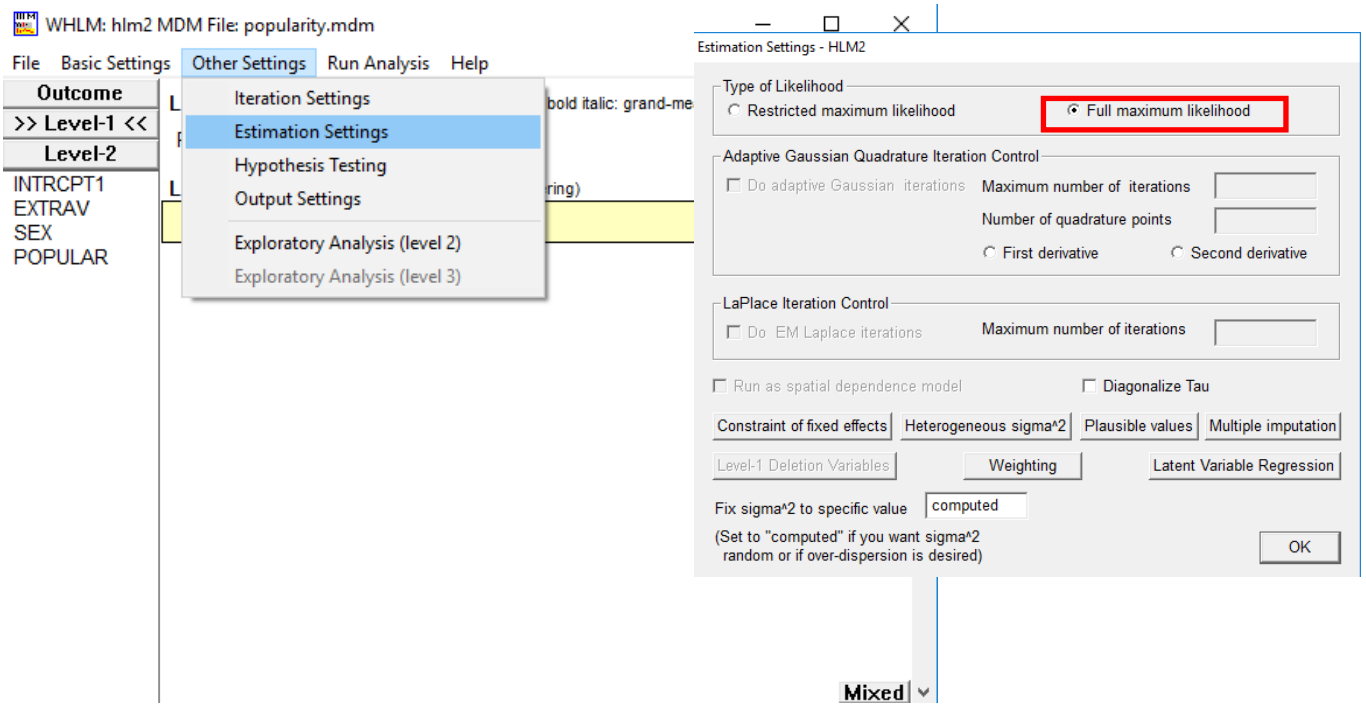
STEP 3: Model properties and options

Now that we have successfully loaded the data into the program, we can start building multilevel models. The first thing that has to be done is specifying the outcome variable. Press **POPULAR** and select **Outcome Variable**.

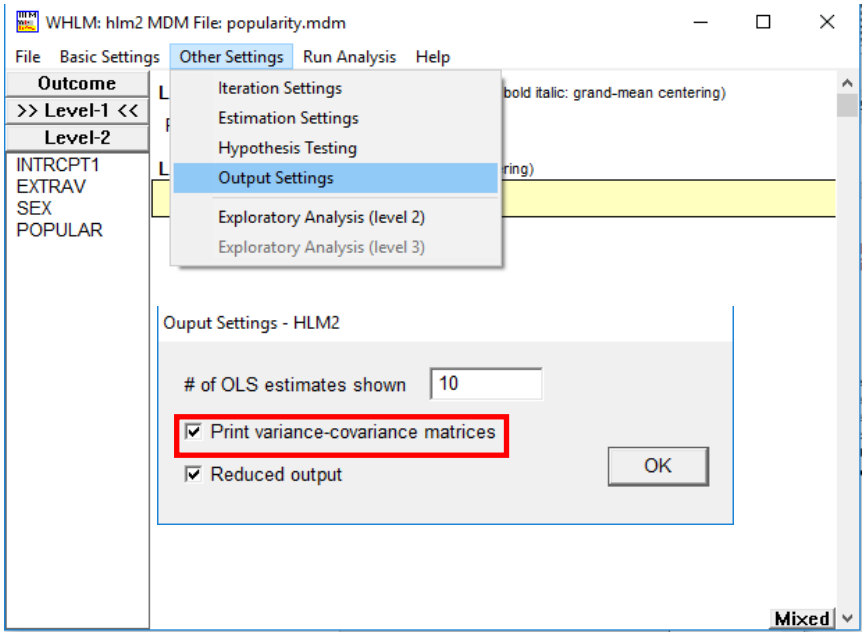




We can now start building models. However, there are a few options we want to change first. In the HLM7 menu bar, go to **Other Settings** > **Estimation Settings**. In the window that pops up, select **Full Maximum Likelihood**, and then press OK.

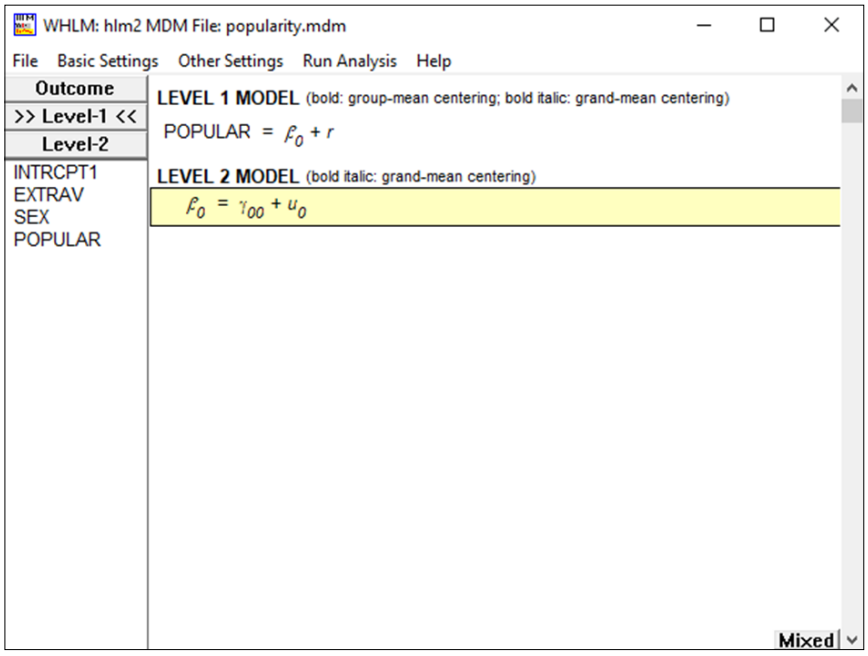


Second, for the final model we will also need the variance-covariance matrix. To ask for this go to **Other Settings** > **Output Settings**. In the window that pops up, select **Print variance-covariance matrix**, and then press OK.

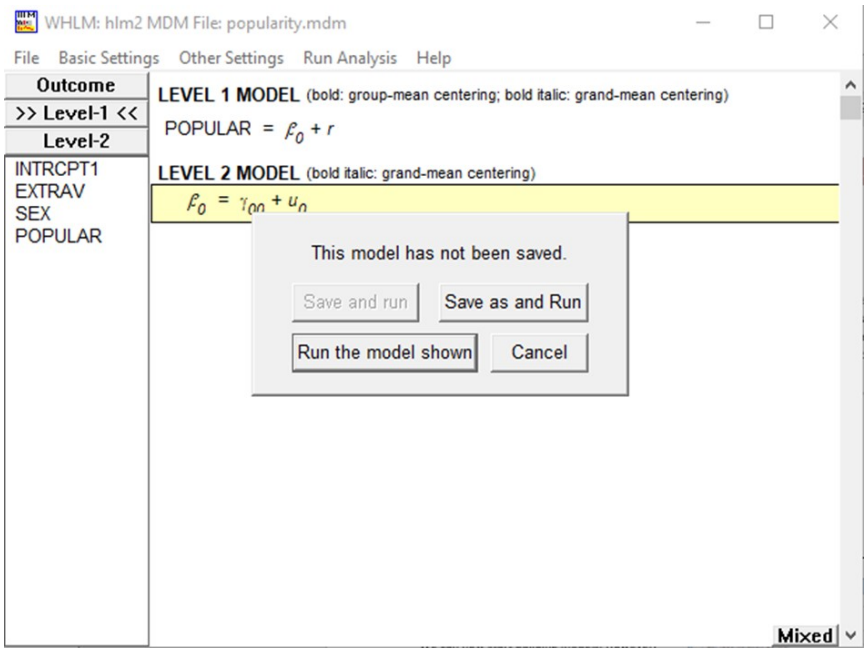


STEP 4: Building Models

Once the data is loaded, and all options have been set, we can start building multilevel models. The first model we want to build is the intercept-only model, with error terms on both levels. Therefore we do not have to add any predictor variables yet. Your window should look like this:



To calculate the model parameters, press **Run Analysis** from the menu bar, and in the window that pops up, press **Run the Model Shown.** You should see the terminal window for a few seconds, which fits the model. After the model has been fitted, a html file with the model output should be saved on the location that you specified during STEP 2.



The HTML file shows all kinds of output from the multilevel analysis. To verify the numbers from the second column Table 2.1 (the first column does not split the variance between levels and is thus not really part of a multilevel analysis) from Hox, Moerbeek, & van de Schoot, scroll down to the end of the text file. The first thing you want to check is usually the **Deviance**. If this seems OK, you can verify the parameter estimates. The fixed effects and variance components should be identical to the ones from Table 2.1. This baseline model without any independent variables can be used to check the intraclass correlation (ICC), and to check whether a multilevel analysis is warranted.

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	5.077859	0.086957	58.395	99	<0.001

Final estimation of variance components

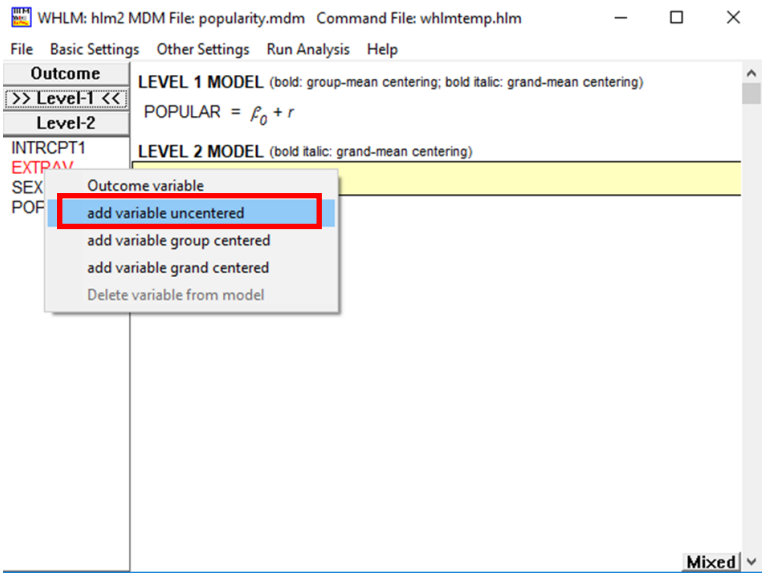
Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	0.83334	0.69446	99	1227.26511	<0.001
level-1, r	1.10535	1.22179			

Statistics for the current model

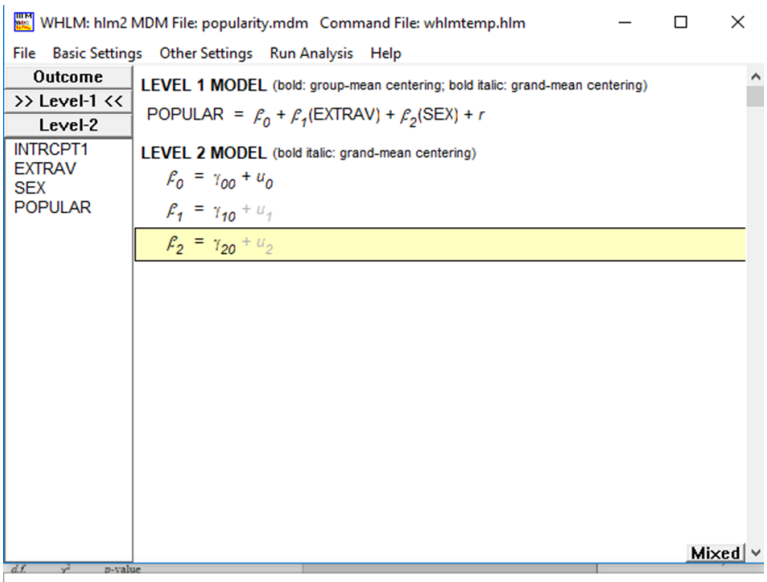
Deviance = 6327.467752
Number of estimated parameters = 3

QUESTION 1: What is the ICC in this example?

We continue by adding predictor variables. To add the level 1 predictors, first click on **Level-1**, then **click the variable you want to add**, and select **Add Variable Uncentered**. The book discusses the advantages of entering variables centered, but in this simple example we will stick to uncentered variables. Do this for EXTRAV and SEX. In the book predictors on both levels are added at the same time. Here we will make the extra step by first only adding the first level variables and later also the second level variables.



The plot on the right shows how your model should look like and on the bottom we see the HTML output. Because we now add two variables, but no interactions yet, we are estimating 5 parameters instead of 3. The results of this output are not given in the book.



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	2.141383	0.116961	18.309	99	<0.001
For EXTRAV slope, β_1					
INTRCPT2, γ_{10}	0.441513	0.016150	27.337	1898	<0.001
For SEX slope, β_2					
INTRCPT2, γ_{20}	1.253137	0.037406	33.501	1898	<0.001

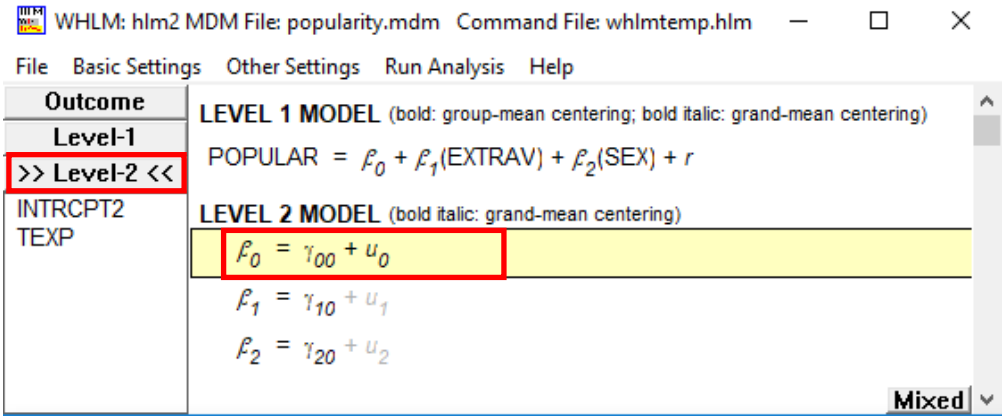
Final estimation of variance components

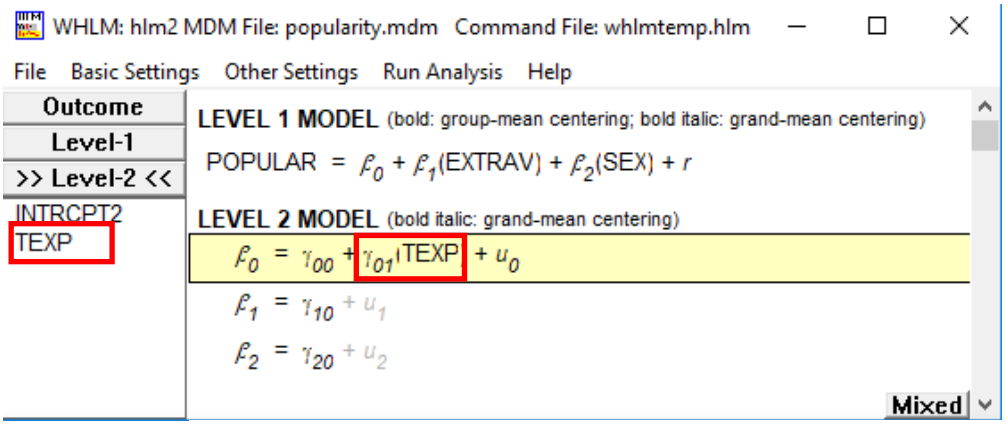
Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	0.78764	0.62037	99	2175.07480	<0.001
level-1, r	0.76906	0.59146			

Statistics for the current model

Deviance = 4933.952263
Number of estimated parameters = 5

We now also (in addition to the level 1 variables that were both significant) add a predictor variable on the second level. **Press Level-2**, and **make sure the β_0 is selected from the Level 2 Model** (the yellow bar). Then **click TEXP** and select **Add Variable Uncentered**, just like you did in level 1.





We can see that both the level 1 and level 2 variables are significant. However, we have not added random slopes yet for any variables (as is done in table 2.1 in the book).

Question 2: What is the explained variance at level 1 and at level 2?

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	0.809326	0.168827	4.794	98	<0.001
TEXP, γ_{01}	0.088409	0.008676	10.190	98	<0.001
For EXTRAV slope, β_1					
INTRCPT2, γ_{10}	0.454484	0.016154	28.134	1898	<0.001
For SEX slope, β_2					
INTRCPT2, γ_{20}	1.254095	0.037265	33.653	1898	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	0.53736	0.28876	98	1072.11886	<0.001
level-1, r	0.76900	0.59136			

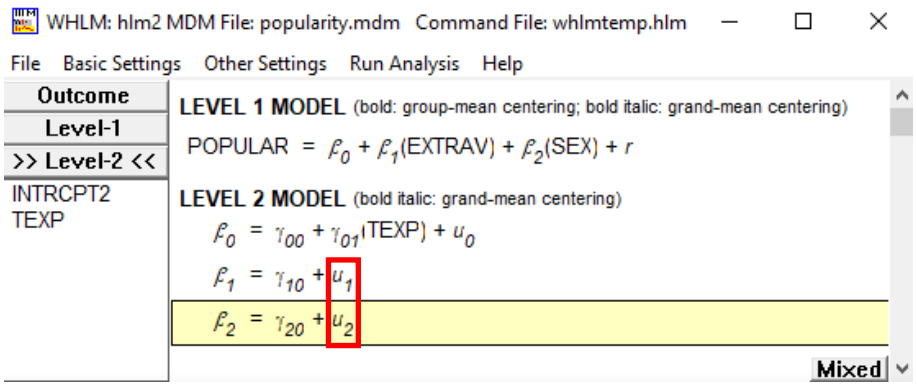
Statistics for the current model

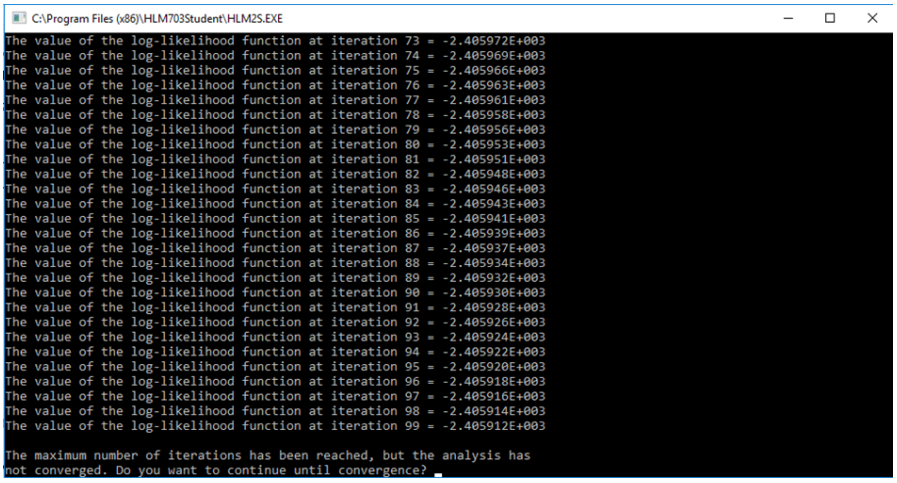
Deviance = 4862.295445
Number of estimated parameters = 6

In the third column of Table 2.1, both predictor variables from level 1 (sex and extraversion) have random slopes. To accomplish this in HLM, activate the error terms for their slopes (β_1 and β_2) by clicking on u_1 and u_2. This will make them turn black instead of grey.

This completes model M1. Press **Run Analysis** like you did before. Now you might encounter a **warning** in the DOS window about convergence. If this happens, type **yes** in the dos window, and press **Enter** to continue fitting the model.

Once done, again open the output file and inspect the parameter estimates like you did before. You might notice that they are slightly different from the ones in Hox, Moerbeek, & Schoot. The reason for this is that the error term for the slope of the variable SEX is very small, and therefore hard to estimate. In the next model this slope will be removed. This model corresponds to the third column of table 2.1 in Hox, Moerbeek, & Schoot. We therefore conclude that there is no slope variation of the SEX variable between classes and therefore the random slope estimation can be dropped from the next analyses.





Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	0.760605	0.195930	3.882	98	<0.001
TEXP, γ_{01}	0.089399	0.008532	10.478	98	<0.001
For EXTRAV slope, β_1					
INTRCPT2, γ_{10}	0.452810	0.024509	18.475	99	<0.001
For SEX slope, β_2					
INTRCPT2, γ_{20}	1.251146	0.037043	33.775	99	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	1.14871	1.31955	86	205.34339	<0.001
EXTRAV slope, u_1	0.18468	0.03411	87	169.19730	<0.001
SEX slope, u_2	0.05670	0.00321	87	80.58867	>0.500
level-1, r	0.74230	0.55101			

Note: The chi-square statistics reported above are based on only 88 of 100 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Statistics for the current model

Deviance = 4811.353882
Number of estimated parameters = 11

$\sigma^2 = 0.55101$

Standard error of $\sigma^2 = 0.01877$

τ

INTRCPT1,β_0	1.31955	-0.18626	-0.02260
EXTRAV,β_1	-0.18626	0.03411	-0.00083
SEX,β_2	-0.02260	-0.00083	0.00321

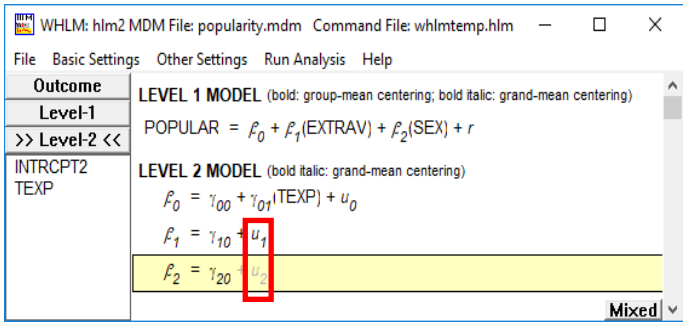
Standard errors of τ

INTRCPT1,β_0	0.28836	0.04691	0.05269
EXTRAV,β_1	0.04691	0.00838	0.00899
SEX,β_2	0.05269	0.00899	0.01861

We continue by clicking on the error term for the parameter of SEX to deselect it (turning from black to grey again), like in the figure on the right.

The estimated variance components and the regression coefficients are now the same as table 2.2 and the 1st column (Model M_1) of table 2.3 in the book.

To get the covariance between the class-level errors for the intercept and extraversion slope we have to look at the off diagonal elements of the variance-covariance output of HLM.



Question 3:

a. What is the intercept of the model?

b. What is the fixed effect of sex?

c. What is the effect of teacher experience?

d. What is the mean effect of extraversion?

e. What is the random effect of the slope of extraversion?

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	0.738568	0.195209	3.783	98	<0.001
TEXP, γ_{01}	0.090813	0.008598	10.563	98	<0.001
For EXTRAV slope, β_1					
INTRCPT2, γ_{10}	0.452641	0.024478	18.491	99	<0.001
For SEX slope, β_2					
INTRCPT2, γ_{20}	1.252470	0.036554	34.263	1799	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	1.13200	1.28143	98	288.64116	<0.001
EXTRAV slope, u_1	0.18417	0.03392	99	237.61747	<0.001
level-1, r	0.74284	0.55181			

Statistics for the current model

Deviance = 4812.806967
Number of estimated parameters = 8

$\sigma^2 = 0.55181$

Standard error of $\sigma^2 = 0.01837$

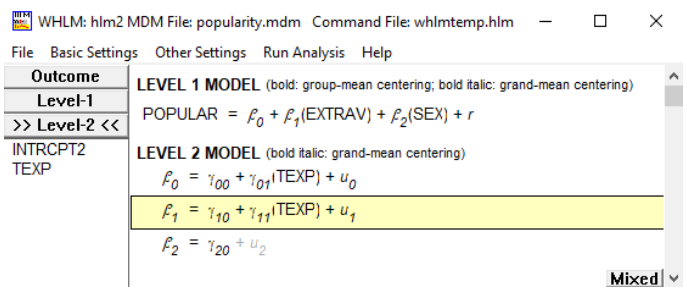
τ

INTRCPT1,β_0	1.28143	-0.18475
EXTRAV,β_1	-0.18475	0.03392

Standard errors of τ

INTRCPT1,β_0	0.28119	0.04655
EXTRAV,β_1	0.04655	0.00833

In the next step to reproduce Model M_2 from Table 2.3, we add the crosslevel interaction between Extraversion and Teacher experience. This means we have to add TEXP as a predictor for the coefficient of EXTRAV. To do this in HLM select below **LEVEL 2 MODEL** the line with the coefficient for EXTRAV. This line should become yellow. Then click TEXP from the variable list on the left, and click **Add Variable Uncentered**. The model should then look like the figure on the right. Run the Analysis and view the output. The Deviance and model parameters are very similar to the ones from model M_2 in table 2.3. You will find the coefficient for the cross-level interaction at the table of fixed effects, below the predictors for EXTRAV. It is somewhere between -0.02 and -0.03, depending on the convergence and rounding settings of the software.



Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-1.207112	0.269006	-4.487	98	<0.001
TEXP, γ_{01}	0.226036	0.016620	13.600	98	<0.001
For EXTRAV slope, β_1					
INTRCPT2, γ_{10}	0.803142	0.039564	20.300	98	<0.001
TEXP, γ_{11}	-0.024699	0.002520	-9.803	98	<0.001
For SEX slope, β_2					
INTRCPT2, γ_{20}	1.240624	0.036203	34.269	1799	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	0.67405	0.45434	98	172.85447	<0.001
EXTRAV slope, u_1	0.06880	0.00473	98	123.39265	0.042
level-1, r	0.74335	0.55256			

Statistics for the current model

Deviance = 4747.624382

$\sigma^2 = 0.55256$

Standard error of $\sigma^2 = 0.01835$

τ

INTRCPT1,β_0	0.45434	-0.02910
EXTRAV,β_1	-0.02910	0.00473

Standard errors of τ

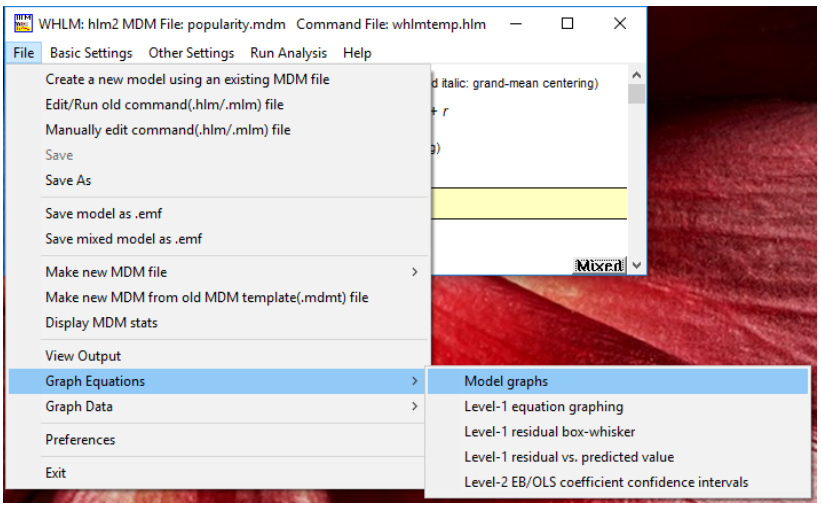
INTRCPT1,β_0	0.15903	0.02353
EXTRAV,β_1	0.02353	0.00399

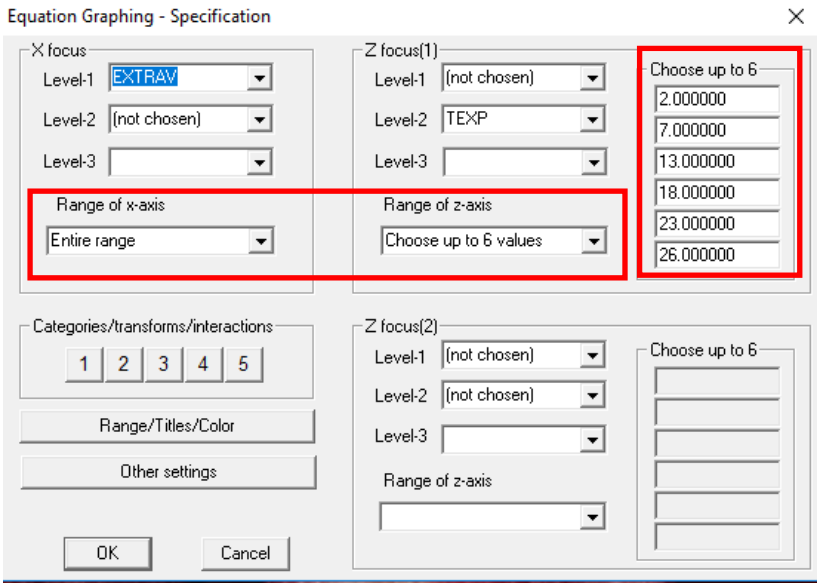
Question 4: What is the explained slope variance of extraversion by using teacher experience as second level variable?

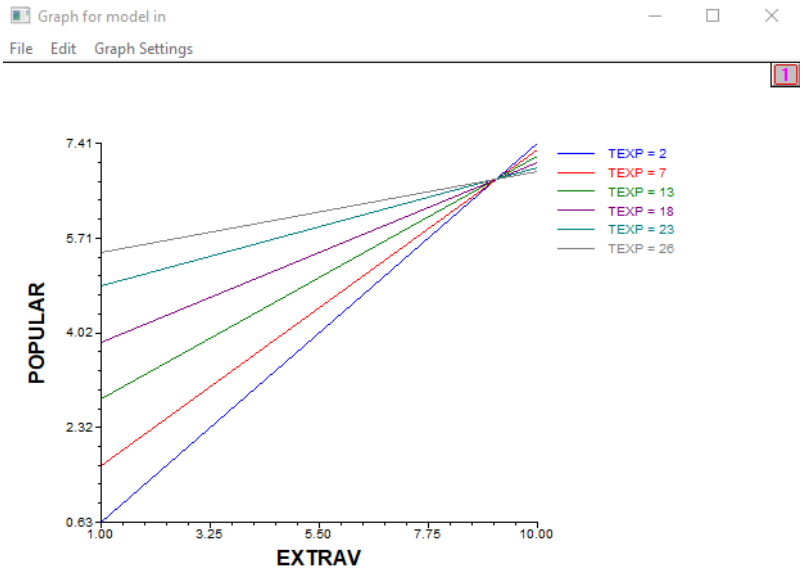
Question 5: What is the number of estimated parameters in this model?

As explained in the book and shown in the results, both the intercept and the slope of the coefficient of extraversion on popularity is influenced by teacher experience. A male student (SEX=0) with a extraversion score of 0 in a class with a teacher with 0 years of experience has an expected popularity of –1.207 (these values are of course impossible, hence the importance of centering). A similar (male) student will improve its popularity with .80 points for every point more extraversion. When teacher experiences increases, the intercept also increases with .23 for every year of experience. So the same male student with no extraversion in a class with a teacher with 15 years of experience has an expected popularity score of –1.207+(15*.226)= 2.183. The teacher experience also lessens the effect of extraversion on popularity. For a teacher with 15 years of experience the regression coefficient of extraversion on popularity is only 0.803-(15*0.0247)=0.4325 (compared to .803 in a class with a teacher with 0 years of experience).

In HLM we can visualize these effects. By clicking **File> Graph Equations > Model graphs .** Then put Extraversion on the X-axis and select the **Entire range** as range of the x-axis. Use TEXP as a level-2 focus and select for the range of the "z"-axis 6 values of your choice (we picked 2 to 26 years of teacher experience).







Question 6: visualize the findings we wound earlier, that there is interecept variance of SEX but no slope variance.

STEP 5: Answers to question:

Question 1: 0.69/(1.22+0.69)=.36

Question 2:

- R2 at level 1: (1.222-0.591)/1.222= .52
- R2 at level 2: (0.694-0.289)/0.694= .58

Question 3:

a. The intercept is 0.738

b. The fixed effect of sex is 1.252

c. The effect of teacher experience is 0.091

d. The mean effect of extraversion is 0.453

e. The random effect of the slope of extraversion is 0.034

Question 4: (0.03392-0.00473)/0.03392= .86

Question 5: 9

Question 6: Two possibilities: