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 Mplus (version 8) for Windows was used. Results should be very similar to results obtained with other software packages, however due to the difference in estimation of the variance/covariance matrices in Mplus and other SEM software compared other statistical modelling software, you might notice minor differences (for more information refer to chapter 14 and 15 of the book). This tutorial is based on earlier versions of similar manuals created by the authors of the book 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 www.github.com/MultiLevelAnalysis/Datasets-third-edition-Multilevel-book/blob/master/chapter%202/popularity/MPLUS/popular2.dat. We will use the .dat file in Mplus, which can be found in the MPLUS folder. This file has already been trans-formed to be able to be used with Mplus.

STEP 2: Loading the data.

To load the .dat datafile and run the analyses it is important that the file is saved in the same folder as the input file for Mplus. When you open Mplus a new input file (.inp file) is automatically opened, you can just work in this file and save your work in the same folder as the popular2.dat file.

STEP 3: Building Models

Once the data is loaded, 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. To do so Mplus requires a few input commands. In order these are:

1.   The **TITLE :** of the file (not very important and not the same as the name of the file)

2.   The **DATA:** This command is used to specify the datafile that is used for the input. Make sure you end every command with a ";", otherwise Mplus will not know you ended command.

3.   The **VARIABLE:** This command is used to specify which variables are in the dataset and which of these variables are being used. Remember that Mplus does require .dat files as an input without variable labels and that you have to name the variables from left to right, just as in the dataset itself. If you mix up the naming the variables in the right order Mplus will not know and your analysis will not yield the right results.

   A.   First you specify the ALL the variables in your dataset, separated by a space after the **NAMES ARE** command. There are quite some variables in the dataset we do not use, but we still need to name them.

   B.   Second, after the **USEVARIABLES ARE** we specify which variable we will use for the analysis

   C.   Third, after the **CLUSTER IS** we specify the variables that indicates the grouping variable (in this example CLASS)

   D.   Fourth, after the **WITHIN ARE** we specify which variables are of the first level (within clusters)

   E.   Fifth, after the **BETWEEN ARE** we specify which variables are of the second level (between clusters)

4.   The **ANALYSIS**: This command is needed to specify what kind of analysis we want to preform

   A.   We want to a two level multilevel model, so under **TYPE IS** we specify **TWOLEVEL** (and later **RANDOM**)

   B.   In line with the book, we can specify the **ESTIMATOR** as Maximum Likelihood (ML), the number of **ITERATIONS =** (for example 1000), and the **CONVERGENCE =** (for example 0.0000001). You could also leave this blank and use the Mplus defaults).

5.   The **MODEL:** this command is needed to specify the model. In a two level multilevel it consists of two part the

   A.   %within%, for variable of the first level

   B.   %between%, for variables of the second level

6.   The **OUTPUT:** This command is needed to specify all kinds of outputs we might want, such as sample statistics, the co-variance matrices, etc.

STEP 3.1: Intercept Only Model (Input)

Based on the description above we can now specify the intercept only model. In this model popular is the only variable we use, so that will be the only variable we specify under USEVARIABLES ARE. Since there are no independent variables yet we leave the WITHIN and BETWEEN ARE commands empty. For the %within% and the %between% part we also indicate that we only have popular as variable at both levels.

Mplus input
```
TITLE:  Popularity data using Mplus, empty model
DATA:
  FILE IS "popular2.dat";
VARIABLE:
  NAMES ARE class pupil cons extrav sex texp popular popteach zextrav
  zsex ztexp zpopular zpoptch;
  USEVARIABLES ARE popular;
  CLUSTER IS class;
  WITHIN ARE ; !empty for now
  BETWEEN ARE ; !empty for now
ANALYSIS:
  TYPE IS TWOLEVEL;
  ESTIMATOR IS ML;
  ITERATIONS =1000;
  CONVERGENCE = 0.0000001;
MODEL:
%within%
popular
%between%
popular
```

OUTPUT:

STEP 3.1: Intercept Only Model (output)

If we then have a look at the output we see that Loglikelihood is –3163.734, which equals the deviance of the intercept only model in the book of the first model (deviance is –2*LogLL).

```
Loglikelihood

        H0 Value                      -3163.734
        H1 Value                      -3163.734
```

If we look at the model results we see that the intercept for popularity is 5.078, that the within (first level) variance is 1.222 and the between level variance 0.695. 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. The ICC is 0.695/(1.222+0.695)=.36.

```
MODEL RESULTS

                                                          Two-Tailed
                    Estimate     S.E.   Est./S.E.   P-Value

Within Level

  Variances
    POPULAR           1.222      0.040    30.822      0.000

Between Level

  Means
    POPULAR           5.078      0.087    58.390      0.000

  Variances
    POPULAR           0.695      0.107     6.489      0.000
```

STEP 3.2: First level variables Only Model (Input)

We continue by adding predictor variables. To add the level 1 predictors, just add them to the model. 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. Since we are now going to use some first level variables we have to include them in the USEVARIABLES ARE command. Since there are now independent variables at the WITHIN level we add them there as well and leave BETWEEN ARE command empty for now. For the %within% and the %between% part we also indicate that we have popular as variable at both levels and add the (regressed) ON command for EXTRAV and SEX at the %within% level. The ON commands mean we regress popularity (dependent variable) on extraversion and SEX (independent variables). In red it is indicated what was added to this input in addition to the previous.

Mplus input
TITLE: Popularity data using Mplus

DATA:
FILE IS "popular2.dat";

VARIABLE:
  NAMES ARE class pupil cons extrav sex texp popular popteach zextrav
  zsex ztexp zpopular zpoptch;
  USEVARIABLES ARE popular extrav sex;
  CLUSTER IS class;
  WITHIN ARE extrav sex;
  BETWEEN ARE ; !empty for now

ANALYSIS:
  TYPE IS TWOLEVEL;
  ESTIMATOR IS ML;
  ITERATIONS =1000;
  CONVERGENCE = 0.0000001;

MODEL:
%within%
popular on extrav;
popular on sex;
%between%
popular

OUTPUT:

STEP 3.2: First level variables Only Model (output)

If we look at the model results we see that the the intercept (MEAN) is now 2.14, the regression coefficient for SEX is 1.25, and the regression coefficient for extraversion (EXTRAV) 0.44. In the last column of e of the output we see the P-values, which indicate all regression coefficients are significantly different from 0. The results of this output are not given in the book.

```
MODEL RESULTS

                                                    Two-Tailed
                        Estimate     S.E.  Est./S.E.   P-Value

Within Level

 POPULAR    ON
    EXTRAV               0.442      0.016     27.228     0.000
    SEX                  1.253      0.037     33.449     0.000

 Residual Variances
    POPULAR              0.591      0.019     30.814     0.000

Between Level

 Means
    POPULAR              2.141      0.117     18.280     0.000

 Variances
    POPULAR              0.620      0.092      6.710     0.000
```

STEP 3.3: First and Second Level Variables Model (Input)

We now also (in addition to the level 1 variables that were both significant) add a predictor variable on the second level (teacher experience). As we are now going to use some second level variables we have to include them in the USEVARIABLES ARE command. Since there are now independent variables at the WITHIN and BETWEEN level, we add them there as well. For the %within% and the %between% part we also indicate which variables we are (regressed) ON command for EXTRAV and SEX at the %within% level and TEXP at the %between% level. The ON commands mean we regress popularity (dependent variable) on extraversion, SEX and TEXP (independent variables). In red it is indicated what was added to this input in addition to the previous.

Mplus input

```
TITLE:  Popularity data using Mplus
DATA:
    FILE IS "popular2.dat";

  VARIABLE:
    NAMES ARE class pupil cons extrav sex texp popular popteach zextrav
    zsex ztexp zpopular zpoptch;
    USEVARIABLES ARE popular extrav sex texp;
    CLUSTER IS class;
    WITHIN ARE extrav sex;
    BETWEEN ARE texp;
  ANALYSIS:
    TYPE IS TWOLEVEL;
    ESTIMATOR IS ML;
    ITERATIONS =1000;
    CONVERGENCE = 0.0000001;

  MODEL:
  %within%
  popular on extrav;
  popular on sex;

  %between%
   popular on texp;

  OUTPUT:
```

STEP 3.3: First and Second Level Variables Model (output)

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).

For extrav the estimate is .454, for sex 1.254, and for teacher experience .088

We can now also calculate the explained variance at level 1 and at level 2., compared to the base model.

For level 1 this is
$(1.222 - 0.591)/1.222 = .516$

For level 2 this is:
$(0.695 - 0.289)/0.695 = .584$

```
MODEL RESULTS

                                                       Two-Tailed
                         Estimate      S.E.  Est./S.E.   P-Value

Within Level

 POPULAR    ON
    EXTRAV              0.454        0.016    28.126      0.000
    SEX                 1.254        0.037    33.595      0.000

 Residual Variances
    POPULAR             0.591        0.019    30.819      0.000

Between Level

 POPULAR    ON
    TEXP                0.088        0.009    10.190      0.000

 Intercepts
    POPULAR             0.809        0.169     4.793      0.000

 Residual Variances
    POPULAR             0.289        0.045     6.393      0.000
```

STEP 3.4: First and Second Level Variables + Random Slopes Model (1) (Input)

Now we also want to include random slopes. In the third column of Table 2.1 of the book, both predictor variables from level 1 (sex and extraversion) have random slopes. To accomplish this in Mplus, we need to tell Mplus first that the type of analysis we want to do is a random slopes analysis, we do this by adding RANDOM after TWOLEVEL in the TYPE IS command. Then, in the %within% model part we create a new random variable we want to estimate by putting it to the left side of a horizontal bar ("|"). Finally, we need to correlate the newly created random slope variables with eachother and with popular. The rest of the input stays the same and in red it is indicated what was added to this input in addition to the previous.

Mplus input

```
TITLE:  Popularity data using Mplus
 DATA:
    FILE IS "popular2.dat";

 VARIABLE:
   NAMES ARE class pupil cons extrav sex texp popular popteach zextrav
   zsex ztexp zpopular zpoptch;
   USEVARIABLES ARE popular extrav sex texp;
   CLUSTER IS class;
   WITHIN ARE extrav sex;
   BETWEEN ARE texp;
 ANALYSIS:
   TYPE IS TWOLEVEL RANDOM;
   ESTIMATOR IS ML;
   ITERATIONS =10000;
   CONVERGENCE = 0.0000001;

 MODEL:
 %within%
 randomextrav | popular on extrav;
 randomsex | popular on sex;

 %between%
 popular on texp;
  randomextrav with popular;
  randomsex with  popular;
  randomextrav with randomsex;

 OUTPUT:
```

STEP 3.4: First and Second Level Variables + Random Slopes Model (1) (output)

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. Only the variation in classes of the regression coefficient of sex is not significantly different from 0, the mean regression coefficient is still significant (1.252). So there is a gender effect, it is just the same for different classes.

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.   P-Value
Within Level

 Residual Variances
    POPULAR          0.550       0.019    29.538     0.000

Between Level

 POPULAR     ON
    TEXP             0.089       0.012     7.454     0.000

 RANDOMEX WITH
    POPULAR         -0.186       0.048    -3.861     0.000
    RANDOMSEX       -0.001       0.011    -0.087     0.931

 RANDOMSE WITH
    POPULAR         -0.024       0.066    -0.362     0.718

 Means
    RANDOMEXTR       0.453       0.025    18.420     0.000
    RANDOMSEX        1.252       0.038    33.067     0.000

 Intercepts
    POPULAR          0.763       0.231     3.309     0.001

 Variances
    RANDOMEXTR       0.034       0.008     4.013     0.000
    RANDOMSEX        0.007       0.023     0.295     0.768

 Residual Variances
    POPULAR          1.320       0.318     4.147     0.000
```

STEP 3.5: First and Second Level Variables + Random Slopes Model (2) (Input)

We continue by deleting the random part of the sex variable, while keep having extraversion as a random variable. The estimated variance components and the regression coefficients are now the same as table 2.2 and the first column (Model M_1) of table 2.3 in the book. In this model we only omit parts compared to the previous model.

Mplus input

```
TITLE:  Popularity data using Mplus
DATA:
FILE IS "popular2.dat";

VARIABLE:
NAMES ARE class pupil cons extrav sex texp popular popteach zextrav zsex ztexp zpopular zpoptch;
USEVARIABLES ARE popular extrav sex texp;
CLUSTER IS class
WITHIN ARE extrav sex;
BETWEEN ARE texp;
ANALYSIS:
   TYPE IS TWOLEVEL RANDOM;
   ESTIMATOR IS ML;
   ITERATIONS =10000;
   CONVERGENCE = 0.0000001;

MODEL:
   %within%
   randomextrav | popular on extrav;
   popular on sex; ! We no longer create randomsex


   %between%
   popular on texp;
   randomextrav with popular;
    ! Here we deleted the now non-necessary correlations

   OUTPUT:
```

STEP 3.5: First and Second Level Variables + Random Slopes Model (2) (output)

MODEL RESULTS

We see that:

- The intercept is 0.738
- The fixed effect of sex is 1.252
- The effect of teacher experience is 0.091
- The mean effect of extraversion is 0.453
- The random effect of the slope of extraversion is 0.034
- First level residual variance is 0.552
- The residual variance on the second level is 1.280

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
POPULAR ON				
SEX	1.252	0.037	34.213	0.000
Residual Variances				
POPULAR	0.552	0.018	30.017	0.000
Between Level				
POPULAR ON				
TEXP	0.091	0.012	7.643	0.000
RANDOMEX WITH				
POPULAR	-0.185	0.048	-3.864	0.000
Means				
RANDOMEXTR	0.453	0.024	18.490	0.000
Intercepts				
POPULAR	0.738	0.228	3.239	0.001
Variances				
RANDOMEXTR	0.034	0.008	4.038	0.000
Residual Variances				
POPULAR	1.280	0.308	4.152	0.000

STEP 3.6: 1st and 2nd Level Predictors + Random Slopes + Crosslevel Interaction (Input)

As a final step we can add a cross-level interaction between teacher experience and extraversion (since this had a significant random effect, that we might be able to explain). In this next step to reproduce Model M2 from Table 2.3, we add the cross-level interaction between extraversion and teacher experience. This means we have to add TEXP as a predictor for the coefficient of EXTRAV. The cross level interaction term between extraversion and teacher experience can be created by regressing random extraversion (in a sense a dependent variable, we want to explain) on the independent variable teacher experience (TEXP). By doing this we hope that TEXP can explain some variance in the random slope of extraversion.

Mplus input

```
TITLE:  Popularity data using Mplus
 DATA:
    FILE IS "popular2.dat";
   VARIABLE:
    NAMES ARE class pupil cons extrav sex texp popular popteach zextrav
    zsex ztexp zpopular zpoptch;
    USEVARIABLES ARE popular extrav sex texp;
    CLUSTER IS class;
    WITHIN ARE extrav sex;
    BETWEEN ARE texp;
   ANALYSIS:
    TYPE IS TWOLEVEL RANDOM;
    ESTIMATOR IS ML;
    ITERATIONS =10000;
    CONVERGENCE = 0.0000001;

   MODEL:
   %within%
   randomextrav | popular on extrav;
   popular on sex;

   %between%
   popular on texp;
   randomextrav with popular;

   randomextrav on texp; !cross level interaction

   OUTPUT:
```

STEP 3.6: 1st and 2nd Level Predictors + Random Slopes + Crosslevel Interaction (output)

We see that the interaction term is denoted under RANDOMEXTR ON TEXP and is estimated at -0.025.

From these results we can now also calculate the explained slope variance of extraversion by using teacher experience as second level variable: (0.034-0.005)/0.034=.85 (These results are slight different than in the book and then in HLM, that is because different estimation and rounding methods are used.)

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 0.803 points for every point more extraversion. When teacher experiences increases, the intercept also increases with 0.226 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 x.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-(15x0.025)=0.428 (compared to 0.803 in a class with a teacher with 0 years of experience).

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate     S.E.   Est./S.E.    P-Value

Within Level

 POPULAR    ON
   SEX            1.241        0.036     34.170       0.000

 Residual Variances
   POPULAR        0.553        0.018     30.002       0.000

Between Level

 RANDOMEXTR ON
   TEXP          -0.025        0.003     -9.780       0.000

 POPULAR    ON
   TEXP           0.226        0.017     13.579       0.000

 RANDOMEX WITH
   POPULAR       -0.029        0.024     -1.196       0.232

 Intercepts
   POPULAR       -1.207        0.269     -4.481       0.000
   RANDOMEXTR     0.803        0.040     20.260       0.000

 Residual Variances
   POPULAR        0.454        0.160      2.828       0.005
   RANDOMEXTR     0.005        0.004      1.103       0.270
```