Multiple Group Multilevel Analysis

Tihomir Asparouhov and Bengt Muthén

Mplus Web Notes: No. 16

November 15, 2012

1 Introduction

Currently a comprehensive multilevel multiple group modeling framework is not well established and well defined. This is particularly true when the group variable is defined on the individual level. Consider the following simple example where students are nested within classrooms/teachers and are grouped by gender. Naturally we want to know if the teachers effect on the student's performance is the same or different across gender. If the teacher effect on the students is different for boys and girls we would want to know if these effects are correlated and to what extent. If the teacher effect on the students is different across gender, are these differences the same across the teachers and could these differences be predicted by the gender of the teacher, the teacher's qualifications or the subject that is taught. A comprehensive modeling framework should be able to address all these questions while accounting further for any group specific differences such as for example different variability within group.

In this note we describe various multiple group modeling possibilities available in Mplus for multilevel data. We discuss the advantages and disadvantages for the different models. Interpretation and comparison for the models are provided as well as the the specific estimation limitations for each model and estimation method. Models with continuous and categorical dependent variables are discussed using three different estimators: ML, Bayes and WLSMV. Input files for Montecarlo data generation and input files for all modeling possibilities are provided in the Appendices at the end of this note. All examples can be replicated by the reader using Mplus Version 7.

We focus on two-level modeling, however, many of the concepts and issues can be extended to three-level modeling as well. Mplus modeling capabilities for three-level multiple group analysis are not as extensive as those for two-level models and thus not all of the illustrations presented here can be easily extended to three-level models.

In single level models a discrete/group variable can affect only the means of the dependent variables or the means and the variance/covariances. We call the first model M_1 and the second model M_2. An example of an M_1 model is a MIMIC model or a CFA model with covariates where the group variable is treated as a covariate. Dummy indicator variables are created for each group, i.e., for each group a variable is created that is 1 if the observation belongs to that group and 0 otherwise. To estimate the M_1 model the dummy variables are used as covariates for the dependent variables. An example of an M_2 model is multiple group CFA model where some parameters are held equal across the groups and some are not held equal. If loadings or variance parameters are group specific then the estimated variance/covariance matrix for the observed variables will be group specific.

In twolevel multiple group modeling, however, there are more variation and complexities because not only can the parameters be different across groups but also the latent variables. More specifically, if a cluster contains observations from several groups the between level cluster variables can be group specific, in particular the between level part of an observed variable can be group specific. This modeling concept occurs only when there are multiple groups in each cluster. For example, if the grouping variable is gender and the cluster variable is school and there are males and females in each or some schools then the between part of an observed variable can be different across the two genders. The interpretation of the between part of a variable in two-level modeling is as usual the average cluster value across all individuals in that cluster. Thus having separate between parts for the two genders simply amounts to accommodating in the model the belief that the average values for males and females is different for all or some of the clusters. This modeling concept is quite natural, however, there are several hurdles that have to be overcome in the process of establishing this. The overall means are different across gender, because the fixed intercept or mean parameters are different across groups. Thus when we are considering the concept of group specific between parts we are modeling cluster specific gender effects that go beyond the overall gender differences that hold for the entire population which is already modeled in model M_1. We discuss these concepts in greater details below and will illustrate with specific Mplus examples.

The group specific between latent variables concept exists only when a clusters contain observations from different groups. If for example the grouping variable is public v.s. private schools then each cluster contains observations only from one group and there is no need to model group specific between level latent variables. Thus in the discussion below we make a clear separation between these two types of grouping variables. The first type is when the grouping variable is a between level discrete variable and all the observations in each cluster belong to exactly one group. The second type is when the grouping variable is a within level discrete variable and the clusters can contain observations from more than one group. The modeling options for a between level group variables are discussed in Section 2 and for a within level group variables are discussed in Section 3.

This note is intended to challenge the reader and stir more questions than to provide all answers. We illustrate some of the modeling and estimation choices available in Mplus. For simplicity we first present the modeling possibilities using only a single dependent variable although the discussion below applies to general multivariate models. Many more combinations and approaches can in principle be constructed for the multivariate case. Models with a single dependent variable are sufficient to illustrate the main modeling concepts, but those concepts can be extended to the multivariate case and they can be extended to latent variables as well. In Section 2 we discuss between level group variable modeling. In Section 3 we discuss within level group variable modeling. In both Section 2 and 3 we use simulated example with one dependent variable. In Section 4 we illustrate some multivariate multilevel multiple group modeling with real data examples for continuous and ordered categorical indicators in a factor analysis model.

2 Between level group variable

An example of a between level group variable is the binary variable indicator for private v.s. public schools when the cluster variable is the school. Another example is the teacher's gender when the cluster variable is the classroom. This is the simplest case of multilevel multiple group analysis. The general model is described as follows. Let Y_{ijg} be the observed variable for individual i in cluster j in group g

$$Y_{ijg} = Y_{Wijg} + Y_{Bjg} \tag{1}$$

$$Y_{Wijg} \sim N(\mu_{1g}, \Sigma_{1g}) \tag{2}$$

$$Y_{Bjg} \sim N(\mu_{2g}, \Sigma_{2g}). \tag{3}$$

The meaning of the last two equations is that the within and the between level components can be structured for each group separately. Consider as an example a two-level model with one factor on both levels

$$Y_{Wijg} = \mu_{1g} + \Lambda_{wg}\eta_{wijg} + \varepsilon_{ijg} \tag{4}$$

$$Y_{Bjg} = \mu_{2g} + \Lambda_{bg}\eta_{bjg} + \varepsilon_{jg}. \tag{5}$$

All of the parameters in the above equations are group specific, including the means, the loadings as well as the residual or factor variances. Certain equality constraints are typically needed for the model to be identified which we will not discuss here. The above model is illustrated in Mplus user's guide example 9.11, Muthén and Muthén (1998-2012). The corresponding Montecarlo and data generation example is included with the Mplus installation. Using the maximum-likelihood estimation in these settings with normally distributed variables does not need numerical integration. The maximum-likelihood estimation is explicit and is based on the EM algorithm. Because the groups are independent of each other the complete log-likelihood is simply the sum of the log-likelihoods across the different groups

$$LL = \sum_{g=1}^{G} LL_g \tag{6}$$

where LL_g is the log-likelihood for group g, LL is the total log-likelihood, and G is the total number of groups. Thus the optimization algorithm is a simple repetition of the optimization algorithm used for the estimation of a single group multilevel model with the added complexities of across groups parameter equalities.

In these settings, the grouping variable is nested above the cluster variable. Thus if the number of groups is 10 or more, then three level modeling should be considered as this will yield a more parsimonious and accurate model. The third level will be the grouping variable. An example for this kind of analysis is Mplus User's Guide example 9.21. Using three level modeling only the means are group specific, i.e., this modeling yields an M_1 model.

It is possible to use the WLSMV estimator in these settings when there are normal and categorical dependent variables, however this estimator does not currently allow multiple group modeling for two-level analysis and thus only the dummy variable approach can be used to estimate M_1 models, see Mplus user's guide example 9.4.

When using the maximum-likelihood estimator with two-level categorical data numerical integration is used. Each observed variable can potentially lead to one dimension of integration because it adds one random intercept. Thus with many categorical outcomes it is important to reduce the number of dimensions of integration by retaining only the most important features in the model. For example, model (5) can be replaced with this model

$$Y_{Bjg} = \mu_{2g} + \Lambda_{bg}\eta_{bjg} \tag{7}$$

if estimation problems occur due to the numerical integration. Model (5) estimation uses P dimensional integration, where P is the number of observed categorical variables, while model (7) estimation uses M dimensional integration, where M is the number of latent variables.

In Appendix A we include an Mplus input file for data generation for a two-level two-group model with one binary variable. The setup for this data generation uses two-level mixture model with two classes. Note here that in order to conduct a simulation study for the two-level two-group model this setup is not sufficient because the class variable is latent instead of observed. To make the class variable observed one has to add to this setup a categorical variable or a nominal variable that is a perfect indicator for the latent class variable, i.e., is equal to the latent class variable. This will essentially convert the latent class variable into an observed grouping variable. In Appendix B we include the input file for analyzing the data generated in Appendix A with a two-level two-group model with a between level group variable. Note here that instead of an observed categorical or nominal perfect class indicator variable we have used the KNOWNCLASS option which is another alternative to convert a latent class variable into an observed grouping variable. This option is not available, however, in Montecarlo studies. Note again that Appendices A and B should be used for the analysis of categorical data. For the analysis of continuous data one should use the simpler User's guide example 9.11.

It is possible to estimate the model in Appendix B also with the Bayes estimator through an entirely different setup. Multiple group or two-level mixture modeling options are not currently available in Mplus for the Bayes estimator in two-level settings and thus we need to use an entirely different setup. The Bayes estimator has the advantage over the ML estimator that it can handle any number of observed and latent variables because it does not use numerical integration. The setup for the Bayes estimator is included in Appendix C. In this setup we create two dummy variables for the two groups d_1 and d_2 and we regress the dependent variable on these dummy variables with random regression slopes S_1 and S_2. These regression slopes are essentially the between parts of the categorical variable for the two groups. Note here that since the between parts are never measured for the same cluster the covariance between the two random slopes is an unidentified parameter which should therefore be fixed to 0 to make the model identified. Note also that with the Bayes estimator such a setup is needed also with continuous variables and not just categorical. Note also that with categorical variables the parameter estimates with the Bayes estimator are computed on a probit scale rather than on the logit scale as in Appendices A and B.

We therefore include the MODEL CONSTRAINTS command in Appendix C to compute the parameters in the logit scale. The threshold parameters here are computed as the means of the random slopes and thus the signs should be reversed as well to match the scale of the parameters computed in Appendix B with the ML estimator.

3 Within level group variable

A model with a within level group variable is generally more difficult to estimate because the fundamental equation (6) does not generally hold. The log-likelihood is not a simple sum of the the log-likelihoods across the different groups because the variables from the different groups are generally not independent. In fact the log-likelihood can not be expressed separately for the different groups because the groups share common variables. Each cluster now contains observations from different groups and the cluster level random effects can be different in all the groups. The general model is again given by equations (1-3), however, in each cluster there are multiple random effects and they can be correlated. Thus we have to replace equation (3) with the following equation

$$Y_{Bj.} \sim N(\mu_{2.}, \Sigma_{2.}) \tag{8}$$

where $Y_{Bj.}$ is a vector that contains all the random effects for all the groups $Y_{Bj.} = (Y_{Bj1}, ..., Y_{BjG})$. The mean vector $\mu_{2.}$ and the variance covariance matrix $\Sigma_{2.}$ give the joint distribution of all the random effects $Y_{Bj.}$ and thus the random effects in cluster j for the different groups can be correlated. Consider for example the case where the grouping variable is gender for students clustered within classrooms and an observed variables Y_{ijg} is influenced by different classroom effects for female and male students. The classroom effect for the two genders are naturally correlated due to the fact that the students are in the same classroom, however the classroom effect need not be the same for the two genders, i.e., the average value for the two genders need not be the same. In this example equation (8) can be explicated by following two equations

$$Y_{Bjg} \sim N(\mu_{2g}, \sigma_{2g}) \tag{9}$$

$$Cov(Y_{Bj1}, Y_{Bj2}) = \rho. \tag{10}$$

Equation (10) clearly shows the difference between modeling within and between level group variables. For between level group variables the random effects are not correlated, i.e., are modeled as independent. The correlation parameter ρ is not an identifiable parameter in the between level group variable case as only one of the two random effects exist.

In Appendix D we include an Mplus input file that can be used to generate data according to the above model. The data generation uses the two-level mixture module of the Mplus program. In this input file we include just the model population command which specifies the parameters that are used for the data generation and we don't specify a model command that would be used for the model estimation. The Mplus program instead assumes a default model specification and uses that model in the estimation. The goal however of this input file is not to estimate a model but only to generate the data. In principle we can specify a model command in this input file which is a simple copy of the model population command, however, this model specification yields a mixture model where the binary grouping variable is unobserved, i.e., a latent grouping variable. This is not our goal since we are interested in observed grouping variables such as gender. If we are interested in conducting a simulation study for this multiple group model rather than a two-level mixture model we have to include as a dependent variable a binary variable that is a perfect indicator for the latent class variable. This way we convert the grouping latent variable into an observed variable. Alternatively a simulation study can be conducted using the input file provided in Appendix D and the external Montecarlo facilities illustrated in example 12.6 in the Mplus user's guide. In Appendix D we include an Mplus input file for data generation for a two-level two-group model with one continuous variable. We use this data to illustrate the alternative modeling possibility in Mplus.

The input file in Appendix D also illustrates how Mplus treats within level groups and how between level random effects have to be specified. Unlike the between level group case discussed in the previous section, here Mplus will not allow the specification of the between level random effect to be done with the name of the observed variable. If the random effect is specified using the same name as the observed variable that will mean that there is only one between level effect that is the same for both groups. Instead we are interested in the model where the random effect is different for the two groups and this is why two latent variables are introduced to represent these between level random effects e_1 and e_2. By specifying zero residual variance for Y and loadings that are 0 or 1 in the two groups we essentially specify a model where Y_b is represented by the different e_g in the different groups. It is also important to note here that the two random effects e_1 and e_2 are correlated during the data generation process.

The Mplus input file that can be used to analyze the data generated in Appendix D is included in Appendix E. With this input file we analyze the data using the exact same model that was used to generate the data, i.e., with this input file we estimate different between level effect for each group. We call this model the H_1 model. Note however that this model estimation uses 2 dimensions of numerical integration. In fact if there are G groups the model would use G dimensions of numerical integration. Note that the numerical integration is used even when the dependent variable is continuous. In principle the likelihood is explicit for this model when the dependent variable is continuous and it could be maximized without numerical integration, however, such an explicit likelihood approach is not implemented in Mplus yet for this model and thus the numerical integration approach is the only one available. If the model has more than one variable there could be G random effects for each variable and that will yield even more dimensions of numerical integration. This model has 7 parameters. Note that the Mplus output reports 8 parameters because it includes the parameter that determines the percentage of observations in each group which is typically not included in a multiple group model so we will not count this parameter here. The 7 parameters are: the two group specific means, the two group specific variances on the within level and the two on the between level as well as the covariance parameter between the two group specific cluster effects. The H_1 model can be described with the following equations

$$Y_{ijg} = \mu_g + \xi_{gj} + \varepsilon_{ijg} \tag{11}$$

where Y_{ijg} is the dependent variable for an observation i belonging to group g and cluster j, μ_g are the intercept parameters,

$$\varepsilon_{ijg} \sim N(0, \theta_g) \tag{12}$$

$$\xi_{jg} \sim N(0, \psi_g) \tag{13}$$

and

$$Cov(\xi_{jg_1}, \xi_{jg_2}) = \rho_{g_1 g_2}. \tag{14}$$

There are several competing models that are not as general and flexible as the H_1 model. These models are more restricted, but have easier and more scalable estimation methods. The models are generally nested in the above model and the likelihood ratio test can be used to test if these more restricted models are sufficiently good fit for the data compared to the H_1 model. Note that for small data sets with small number of clusters the LRT test may accept the simpler models as sufficient only because of the lack of power to establish the need for the more general H_1 model. Practical applications of this methodology are needed to determine the performance of the LRT test with real data.

The first alternative to model H_1 is model H_2. This model is estimated with the input file included in Appendix F. In this model the different groups have different cluster level effects, however the correlation between these effects is fixed to 1 and thus there is only one continuous latent dimension and the estimation uses only 1 dimension of numerical integration. The cluster specific deviation from the fixed effect of the group variable on the dependent variable are proportional. Since in many applications the cluster effects will be highly correlated, the assumption of correlation 1 between the cluster level effects maybe reasonable in many real data examples. The model uses one latent variable on the between level while the loadings vary across the groups to provide group specific variance for the cluster effect. Model H_2 has 6 parameters. The only parameter that is missing here is the correlation parameter on the between level which is essentially fixed to 1. This model can be estimated with 1 dimensional numerical integration regardless of how many groups there are. This is a computational advantage over model H_1. The H_2 model estimation uses numerical integration even when the dependent variable is continuous, just like the H_1 model estimation. The H_2 model can be described with the following equations

$$Y_{ijg} = \mu_g + \lambda_g \xi_j + \varepsilon_{ijg} \tag{15}$$

where μ_g are the intercept parameters, λ_g are loading parameters

$$\varepsilon_{ijg} \sim N(0, \theta_g) \tag{16}$$

and

$$\xi_j \sim N(0, 1). \tag{17}$$

Here λ_g^2 can be interpreted as the variance of the group specific cluster effect.

Model H_3 is estimated with the input file included in Appendix G. In this model the correlation between the two between level random effects is fixed to 0 instead of 1. This model does not require numerical integration at all and is estimated with the non-mixture module of Mplus using standard multiple group analysis. If the random effects are independent across groups then we can split each cluster into two independent clusters each consisting of the observations in the two groups. This is achieved simply by redefining the cluster variable so that the cluster variable has different values in the different groups. This method could perform quite poorly in practice because we can assume in general that the random effects will be correlated. Note also that in this model the likelihood does not include the likelihood for the grouping variable and also does not include as a parameter the proportion of observations in each group. Thus in order to conduct the LRT test between model H_1 and model H_3 for example the likelihood of the grouping variable has to be computed separately and added to the likelihood of the H_3 model. The likelihood of the grouping variable is simply the likelihood for a binary variable in this case. To estimate model H_3 we do not need to specify a model statement because the default model statement in Mplus is the H_3 model. This model also has 6 parameters. The advantage of this model is again a computational advantage. The model does not require any numerical integration regardless of how many variables are used. To estimate this model with a categorical variable however numerical integration will still be used and the estimation should be done through the twolevel mixture module of Mplus. The H_3 model can be described with the following equations

$$Y_{ijg} = \mu_g + \xi_{gj} + \varepsilon_{ijg} \tag{18}$$

where μ_g are the intercept parameters,

$$\varepsilon_{ijg} \sim N(0, \theta_g) \tag{19}$$

and

$$\xi_{jg} \sim N(0, \psi_g). \tag{20}$$

The cluster effects ξ_{jg} are not correlated within cluster j.

Model H_4 is another slightly restricted model. The model is equivalent to model H_1 with the restriction that the within level residual variances are the same across the two groups. The model setup is included in Appendix H. The model doesn't use multiple group or mixture setup. Instead as in Appendix C the model uses dummy variables to create the cluster random effects as random slopes for the dummy variables. The model can be estimated with the ML or Bayes estimator with categorical or continuous variables. The model also has 6 parameters. The only missing parameter is the group specific residual variance on the within level. Here again the log-likelihood does not include the log-likelihood of the grouping variable so to conduct the LRT between model H_1 and this model the log-likelihood has to be adjusted by the grouping variable log-likelihood. The H_4 model can be described with the following equations

$$Y_{ijg} = \mu_g + \xi_{gj} + \varepsilon_{ijg} \tag{21}$$

where μ_g are the intercept parameters,

$$\varepsilon_{ijg} \sim N(0, \theta) \tag{22}$$

$$\xi_{jg} \sim N(0, \psi_g) \tag{23}$$

and

$$Cov(\xi_{jg_1}, \xi_{jg_2}) = \rho_{g_1 g_2}. \tag{24}$$

Model H_5 is another slightly restricted model. The model is a restriction of model H_4. The new restriction here is that the residual variance on the between level is also not a group specific parameter. This model is estimated as a three-level model where the grouping variable is transformed into a new clustering variable. The original cluster variable is now the cluster level variable at the highest level. The multiple groups within the original clusters now represent the second level clustering. The input file for this model is included in Appendix I. The groups specific means are retained via a regression on the grouping variable, or in case of more than two groups on the dummy variables. The group specific cluster effect is also retained in this model as these are the effects from the middle clustering level and they are different across the groups. The group specific cluster effects are also correlated through the level 3 clustering effect. This model has the advantage over the H_4 model in that it can handle more elegantly larger number of groups, by retaining the group specific cluster effects without increasing the number of parameters dramatically as the number of groups increase. The model can be estimated with the ML or Bayes estimators with categorical or continuous variables. The H_5 model can be described with the following equations

$$Y_{ijg} = \mu_g + \zeta_j + \xi_{gj} + \varepsilon_{ijg} \tag{25}$$

where μ_g are the intercept parameters,

$$\varepsilon_{ijg} \sim N(0, \theta) \tag{26}$$

$$\xi_{jg} \sim N(0, \psi) \tag{27}$$

and

$$\zeta_j \sim N(0, \sigma). \tag{28}$$

Models H_6 and H_7 are models where the cluster specific effect is the same across the groups. These models are simpler to estimate and should be considered as well because in many practical situations even if the cluster effects are group specific due to small sample size there will be not enough power in the data to establish statistical significance for the group specific cluster effect models H_1,...,H_5. Model H_6 is nested within model H_2 and is based on the restriction that the cluster effects are not just proportional but actually equal. This is equivalent to saying that the variances of the cluster specific effect are equal across the groups and since the correlation between these effects is 1 then the effects are actually equal. In this model H_6 the group variable has a fixed effect on the mean of the variable and the within level residual varies across the groups. The model has 5 parameters. The parameters are the same as in model H_2 except that the between level variances are the same across the two groups. Model H_6 uses 1-dimensional integration regardless of the type of dependent variable in the models. The H_6 model can be described with the following equations

$$Y_{ijg} = \mu_g + \xi_{jg} + \varepsilon_{ijg} \tag{29}$$

where μ_g are the intercept parameters,

$$\varepsilon_{ijg} \sim N(0, \theta_g) \tag{30}$$

and

$$\xi_{jg} \sim N(0, \psi). \tag{31}$$

Finally model H_7 is a further restriction of model H_6 that holds the within level variances the same across the groups. This model is essentially equivalent to including the grouping variable as a predictor and it does not use Mplus multiple group utilities. The model has just 4 parameters, the two group specific mean parameters as well as the within and the between variance parameters. Model H_7 does not use numerical integration if the dependent variable is continuous and it uses 1-dimensional integration if the variable is categorical. The H_7 model can be described with the following equations

$$Y_{ijg} = \mu_g + \xi_{jg} + \varepsilon_{ijg} \tag{32}$$

Table 1: Log-likelihood values for different model using simulated data

Model	Log-likelihood	Number of parameters
H_1	-20800.8	7
H_2	-20948.4	6
H_3	-20842.4	6
H_4	-21009.7	6
H_5	-21011.2	5
H_6	-21009.5	5
H_7	-21200.2	4

where μ_g are the intercept parameters,

$$\varepsilon_{ijg} \sim N(0, \theta) \tag{33}$$

and

$$\xi_{jg} \sim N(0, \psi). \tag{34}$$

Table 1 contains the likelihood values set in the same metric by including the log-likelihood of the grouping variable and the number of parameters, excluding the group proportion parameter. Since the data was generated according to the most flexible model H_1 it is no surprise that the best likelihood is obtained for this model and a formal LRT test will reject all other model.

In the settings of within level grouping variable, the grouping variable is not nested above the cluster variable but it provides a cross nesting for the observations. Thus if the number of groups is 10 or more, then a cross-classified model should be considered as this will yield a more parsimonious and accurate model. The second clustering variable for the cross classified model will be the grouping variable. Using cross-classified modeling only the means will be group specific.

4 Two-level two-group factor analysis models

In this section we illustrate the multilevel multiple group modeling described in Section 3 using real data examples. In Section 4.1 we illustrate the modeling and the estimation issues for a factor analysis model with continuous indicators and in Section 4.2 we use a factor analysis model with categorical indicators. The type of indicator variables determines the estimation methods that are available, the numerical issues that can be encountered, the dimensions of numerical integration, and the possible and feasible modeling extensions. Thus we consider the case of continuous and categorical variables separately.

4.1 Two-level two-group models with continuous variables

In this section we illustrate the multilevel multiple group modeling using a 1-factor analysis model measured by continuous variables. The multiple group modeling is applied to the factor only. We do not model in these analysis any item specific group effects.

The data comes from the National Educational Longitudinal Study (NELS). We analyze the data as described in Muthén et al. (1997). The sample contains data for 5198 students from 235 schools. The variables that are analyzed are created from testlets covering reading, math, science and history. Sixteen achievement variables are created for each student. The grouping variable is the student's gender and the clustering variable is the school variable. The log-likelihood for the seven models described in Section 3 are given in Table 2. We adjust the log-likelihood so that it includes the log-likelihood for the grouping variable but we exclude the parameter for the group proportions. It is clear from the results in Table 2 that the LRT testing rejects all models but the most flexible model H_1 where the cluster effect is group specific. Note however that in a practical situation other factors may be important in selecting the best model. One such factor is the ability of a model to be generalized easily and elaborated upon. For example if we want to include item specific cluster effects some of the models will need many dimensions of numerical integrations while others will not. Similarly if more latent variables are involved some of these models will result in high dimensional numerical integration while others will not.

The Mplus input files for Models H_1-H_7 are included in Appendices L-R. In model H_3 we used the define statement to split the clusters in two clusters by adding gender*3000000. The value 3000000 is bigger than the maximum cluster value in the original data file and with that statement we are guaranteed that when gender=0 all cluster values are below 3000000 and when

Table 2: Log-likelihood values for different model using NELS data

Model	Log-likelihood	Number of parameters	Dimensions of integration
H_1	-128519.0	53	2
H_2	-128525.1	52	1
H_3	-128607.7	52	0
H_4	-128524.5	52	0
H_5	-128524.8	51	0
H_6	-128525.1	51	1
H_7	-128530.9	50	0

gender=1 all values are above 3000000, i.e., each clusters is split by gender to form two new clusters. In models H_3, H_5, and H_7 we used a regression on the constant 1 to obtain the random intercepts of the factor on the between level(s). To include a constant in the model the variance=nocheck; data command option is needed. An alternative method for estimating this model is to use the approach used for model H_1 but to fix the residual variances of the observed variables to 0. When numerical integration is not used the residual variances are actually fixed to a small value of 0.0001 rather than zero to avoid singularity of the between level variance covariance matrices. In most applications this will work well. In some applications however that approach can lead to slow convergence and imprecision in the log-likelihood value which is the main interest here. In all other examples we hold the loadings parameter on the within and the between level equal so that the between factor can be interpreted as the between part / cluster effect of the main factor.

4.2 Two-level two-group models with categorical variables

In this section we illustrate the multilevel multiple group modeling using a 1-factor analysis model measured by ordered categorical items. Here the multiple group modeling is applied to the factor just as in the previous section and we do not model any item specific group effects.

The data comes from the Johns Hopkins Center for Prevention and Early

Intervention Cohort described in Ialongo et al. (1999). We use Cohort 3 of the TOCA (Teacher Observation of Classroom Adaptation) data. The data consists of a teacher-rated measurement instrument capturing aggressive-disruptive behavior among a sample of U.S. students in Baltimore public schools. The instrument consists of 13 items scored as 1 (almost never) through 6 (almost always). A total of 678 students are observed in 27 classrooms from Fall of Grade 1 through Grade 6 for a total of 8 time points. The observations are nested within classroom/teacher and we use the gender variable as a within-level grouping variable and the classroom variable as the two-level cluster variable. In this analysis we use only the data collected in the Fall of Grade 1. Using the above models we can understand to what extent the aggressive-disruptive behavior in students depends on the gender of the student and the classroom they belong to. The log-likelihood for the seven models described in Section 3 are given in Table 3. Again we adjust the log-likelihood so that it includes the log-likelihood for the grouping variable but we exclude the parameter for the group proportions. It is clear from the results in Table 2 that the LRT testing rejects all models but the simplest model H_7 where the gender variable has a fixed effect on the factor. It is also clear from these results the potential drawback of Model H_3. It is much more realistic assumption in this data to assume that the cluster effect is the same for both genders than to assume that the effects are two independent variables. The dependence between potential group specific cluster effects is critical. The fact that no statistically significant differences between the cluster effects for the different groups was found may reflect the fact that the sample size is small and there is not enough power in this data to detect such difference or that gender differences in behavior are quite uniform across the population, particularly in the early grades.

Another model that is of interest here is a variation of Model H_5. We call this model H_{5a}. In this model we include item specific classroom effects as well as group specific item classroom effects. This model is not nested within any of the models H_1-H_7. The model is estimated with the Bayes estimator and the interesting finding here is that while the group specific cluster effect is not significant the group specific item cluster effects are all significant.

All Models H_1-H_7 as well as model H_{5a} are included in Appendices S-Z. Some of the technical options in the H_1 input are needed only for this data set to resolve the singularity between the random effects since the correlation between these random effects converges to 1. In model H_5 we used a regression on the constant 1 to obtain the random intercepts of the factor on

Table 3: Log-likelihood values for different model using TOCA data

Model	Log-likelihood	Number of parameters	Dimensions of integration
H_1	-6692.9	83	3
H_2	-6692.9	82	2
H_3	-6741.3	82	2
H_4	-6693.0	82	3
H_5	Bayes	81	Bayes
H_6	-6693.6	81	2
H_7	-6693.6	80	2

level 2 and level 3. To include a constant in the model the variance=nocheck; data command option is needed. An alternative method for estimating this model is to use the input for model H_{5a} but to fix the residual variances of the observed variables to 0. This approach however will yield a suboptimal estimation and slower mixing due to singularity matrices on the between level. In all other examples we hold the loadings parameter on the within and the between level equal so that the between factor can be interpreted as the between part / cluster effect of the main factor.

References

[1] Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression and antisocial behavior. American Journal of Community Psychology, 27, 599-642.

[2] Muthén, L.K. and Muthén, B.O. (1998-2012). Mplus Users Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén

[3] Muthén, B., Khoo, S.T. & Gustafsson, J.E. (1997). Multilevel latent variable modeling in multiple populations. Unpublished technical report. http://pages.gseis.ucla.edu/faculty/muthen/articles/Article_074.pdf

5 Appendix A: Input file for data generation for a two-level two-group categorical data

```
montecarlo:
names are u;
nobservations = 10000;
ncsizes = 1;
csizes = 1000(10);
genclasses = cb(2 b);
classes = cb(2);
generate = u(1);
categorical = u;
between = cb;
save=ex1.dat;

analysis: type = twolevel mixture;

model population:

%within%
%overall%

%between%
%overall%
[cb#1*0];

%cb#1%
[u$1*-1];
u*1.1;

%cb#2%
[u$1*0.5];
u*0.6;
```

6 Appendix B: Input file for estimating a two-level two-group model with categorical data with the ML estimator

```
variable:
names=u g cluster;
cluster=cluster;
classes = cb(2);
categorical = u;
between = cb;
knownclass = cb (g = 1 2);

data: file=ex1.dat;

analysis: type = twolevel mixture;

model:

%within%
%overall%

%between%
%overall%
[cb#1*0];

%cb#1%
[u$1*-1];
u*1.1;

%cb#2%
[u$1*0.5];
u*0.6;
```

7 Appendix C: Input file for estimating a two-level two-group model with categorical data with the Bayes estimator

```
variable:
names are u g cluster;
cluster=cluster;
categorical = u;
within=u d1 d2;
usevar=u d1 d2;

define: d1=2-g; d2=g-1;

data: file=ex1.dat;

analysis: type = twolevel random; estimator=bayes;
biter = (1000); proc=2;

model:

%within%
[u$1@0];
s1 | u on d1;
s2 | u on d2;

%between%
[s1*1] (t1);
s1*1.1 (v1);
[s2*-0.5] (t2);
s2*0.6 (v2);
s1 with s2@0;

model constraints:
new(nt1 nt2 nv1 nv2);
nt1=-t1*1.81; nt2=-t2*1.81;
nv1=v1*3.29; nv2=v2*3.29;
```

8 Appendix D: Input file for data generation for a two-level two-group data with within level group

```
montecarlo:
names are y;
nobservations = 10000;
ncsizes = 1;
csizes = 500(20);
genclasses = c(2);
classes = c(2);
save=ex2.dat;

analysis: type = twolevel mixture;

model population:

%within%
%overall%
[c#1*0];

%c#1%
y*1.1;

%c#2%
y*0.6;

%between%
%overall%
e1 by y@0; e2 by y@0; [e1@0 e2@0];
y@0; e1*0.3; e2*0.4; e1 with e2*0.2;

%c#1%
[y*-1]; e1 by y@1;

%c#2%
[y*0.5]; e2 by y@1;
```

9 Appendix E: Input file for estimating a two-level two-group model with within level grouping variable. Model H1.

```
variable:
names are y g cl;
classes = c(2);
knownclass = c (g = 1 2);
cluster=cl;

data: file=ex2.dat;
analysis: type = twolevel mixture;

model:

%within%
%overall%
[c#1*0];

%c#1%
y*1.1;

%c#2%
y*0.6;

%between%
%overall%
e1 by y@0; e2 by y@0; y@0; [e1@0 e2@0];
e1*0.3; e2*0.4; e1 with e2*0.2;

%c#1%
[y*-1]; e1 by y@1;

%c#2%
[y*0.5]; e2 by y@1;
```

10 Appendix F: Input file for estimating a two-level two-group model with within level grouping variable and proportional effects. Model H2.

```
variable:
names are y g cl;
classes = c(2);
knownclass = c (g = 1 2);
cluster=cl;

data: file=ex2.dat;

analysis: type = twolevel mixture;

model:

%within%
%overall%
[c#1*0];

%c#1%
y*1.1;

%c#2%
y*0.6;

%between%
%overall%
e by y; y@0; e@1; [e@0];

%c#1%
[y*-1]; e by y*1;

%c#2%
[y*0.5]; e by y*1;
```

11 Appendix G: Input file for estimating a two-level two-group model with within level grouping variable and independent effects. Model H3.

```
variable:
names are y g cl;
grouping = g (1=g1 2=g2);
cluster=cl;

define: cl=cl+1000*g;

data: file=ex2.dat;

analysis: type = twolevel;
```

12 Appendix H: Input file for estimating a two-level two-group model with within level grouping variable and no group specific within level residual variances. Model H4.

```
variable:
names are y g cl;
within=y d1 d2;
usevar=y d1 d2;
cluster=cl;

data: file=ex2.dat;

define: d1=2-g; d2=g-1;

analysis: type = twolevel random;

model:

%within%
y*1; [y@0];
s1 | y on d1;
s2 | y on d2;

%between%
[s1*-1];
[s2*-0.5];
s1*0.3;
s2*0.4;
s1 with s2*0.2;
```

13 Appendix I: Input file for estimating a two-level two-group model with within level grouping variable and no group specific within or between level residual variances. Model H5.

```
variable:
names are y g cl;
usevar are y g cl cl2;
cluster=cl cl2;
within=g;

define: cl2=cl+1000*g;
g=g-1;

data: file=ex2.dat;

analysis: type = threelevel;

model:

%within%
y on g;

%between cl2%
y;

%between cl%
y;
```

14 Appendix J: Input file for estimating a two-level two-group model with group invariant cluster effects and group specific mean and within level variance. Model H6.

```
variable:
names are y g cl;
classes = c(2);
knownclass = c (g = 1 2);
cluster=cl;

data: file=ex2.dat;

analysis: type = twolevel mixture;

model:

%within%
%overall%
[c#1*0];

%c#1%
y*1.1;
%c#2%
y*0.6;

%between%
%overall%
y*1;

%c#1%
[y*-1];
%c#2%
[y*0.5];
```

15 Appendix K: Input file for estimating a two-level two-group model with group invariant cluster effects and within level variance and group specific mean. Model H7.

```
variable:
names are y g cl;
cluster=cl;
within=g;

define: g=g-1;

data: file=ex2.dat;

analysis: type = twolevel;

model:

%within%
y*1;
y on g;

%between%
y*1;
```

16 Appendix L: NELS Model H1.

```
variable:
names = ....
cluster is cluster;
usevariables = y1-y16;
classes = c(2);
knownclass = c (gender = 0 1);

analysis:
type = twolevel mixture;
process = 8;
estimator = ml;

model:
%within%
%overall%
gw by y1@1
y2-y16* (lam2-lam16);
%c#1%
gw*1;
%c#2%
gw*1;
%between%
%overall%
gb by y1@1
y2-y16* (lam2-lam16);
e1 by gb@0; e2 by gb@0; gb@0;
[e1@0 e2@0];
e1*1; e2*1; e1 with e2*0;
%c#1%
[gb@0]; e1 by gb@1;
%c#2%
[gb*0]; e2 by gb@1;
```

17 Appendix M: NELS Model H2.

```
variable:
names = ....
cluster is cluster;
usevariables = y1-y16;
classes = c(2);
knownclass = c (gender = 0 1);

analysis:
type = twolevel mixture;
process = 8;
estimator = ml;

model:
%within%
%overall%
gw by y1@1
y2-y16* (lam2-lam16);
%c#1%
gw*1;
%c#2%
gw*1;
%between%
%overall%
gb by y1@1
y2-y16* (lam2-lam16);
gb@0;
e by gb; e@1; [e@0];
%c#1%
[gb@0]; e by gb*1;
%c#2%
[gb*0]; e by gb*1;
```

18 Appendix N: NELS Model H3.

```
variable:
names = ....
cluster is cluster;
usevariables = y1-y16 one;
grouping is gender (0 = g1 1 = g2);
within=y1-y16 one;

data: file= ....; variance=nocheck;

define: cluster=cluster+3000000*gender; one=1;

analysis:
type = twolevel random;
estimator = ml;

model:
%within%
gw by y1@1
y2-y16* (lam2-lam16);
y1-y16 (t1-t16);
gb | gw on one;
%between%
[gb@0]; gb*1;

model g2:
%within%
gw*1;
%between%
[gb*0]; gb*1;
```

19 Appendix O: NELS Model H4.

```
variable:
names = ....
cluster is cluster;
usevariables = y1-y16 gender gender2;
within = y1-y16 gender gender2;

analysis:
type = twolevel random;
estimator = ml;

define:gender2=1-gender;

model:
%within%
gw by y1@1
y2-y16* (lam2-lam16);
gw*1; [gw@0];
s1 | gw on gender;
s2 | gw on gender2;
%between%
[s1@0];
[s2*1];
s1*1;
s2*1;
s1 with s2;
```

20 Appendix P: NELS Model H5.

```
variable:
names = ....
cluster is cluster cl2;
usevariables = y1-y16 gender cl2 one;
within= gender y1-y16 one;

data: file= ....; variance=nocheck;

define: cl2=cluster+3000000*gender; one=1;

analysis:
type = threelevel random;
estimator = ml;

model:
%within%
gw by y1@1
y2-y16* (lam2-lam16);
gw*1; [gw@0];
gw on gender;
s | gw on one;
%between cl2%
s*1;
%between CLUSTER%
s*1; [s@0];
```

21 Appendix Q: NELS Model H6.

```
variable:
names = ....
cluster is cluster;
usevariables = y1-y16;
classes = c(2);
knownclass = c (gender = 0 1);

analysis:
type = twolevel mixture;
process = 8;
estimator = ml;

model:
%within%
%overall%
gw by y1@1
y2-y16* (lam2-lam16);
%c#1%
gw*1;
%c#2%
gw*1;
%between%
%overall%
gb by y1@1
y2-y16* (lam2-lam16);
gb*1;
%c#1%
[gb@0];
%c#2%
[gb*0];
```

22 Appendix R: NELS Model H7.

```
variable:
names = ....
cluster is cluster;
usevariables = y1-y16 gender one;
within=gender one y1-y16;

data: file= ....; variance=nocheck;

define: one=1;

analysis:
type = twolevel random;
estimator = ml;

model:
%within%
gw by y1@1
y2-y16* (lam2-lam16);
gw*1;
gw on gender;
gb | gw on one;
%between%
gb*1; [gb@0];
```

23 Appendix S: TOCA Model H1.

```
variable:
names = ....
usevariables = u1-u13;
categorical = u1-u13;
cluster = sgsf93;
classes = c(2);
knownclass = c (gender = 0 1);

analysis:
type = twolevel mixture;
process = 8;
estimator = ml; cholesky=off;
variance=0; mconv=0.1;

model:
%overall%
gw by u1@1
u2-u13* (lam2-lam13);
%c#1%
gw*1;
%c#2%
gw*1;
%between%
%overall%
gb by u1@1
u2-u13* (lam2-lam13);
e1 by gb@0; e2 by gb@0; gb@0;
[e1@0 e2@0];
e1*1; e2*1; e1 with e2*0;
%c#1%
[gb@0]; e1 by gb@1;
%c#2%
[gb*0]; e2 by gb@1;
```

24 Appendix T: TOCA Model H2.

```
variable:
names = ....
usevariables = u1-u13;
categorical = u1-u13;
cluster = sgsf93;
classes = c(2);
knownclass = c (gender = 0 1);

analysis:
type = twolevel mixture;
process = 2;
estimator = ml;

model:
%within%
%overall%
gw by u1@1
u2-u13* (lam2-lam13);
%c#1%
gw*1;
%c#2%
gw*1;
%between%
%overall%
gb by u1@1
u2-u13* (lam2-lam13);
gb@0;
e by gb; e@1; [e@0];
%c#1%
[gb@0]; e by gb*1;
%c#2%
[gb*0]; e by gb*1;
```

25 Appendix U: TOCA Model H3.

```
variable:
names = ....
usevariables = u1-u13;
categorical = u1-u13;
cluster = sgsf93;
classes = c(2);
knownclass = c (gender = 0 1);
between = c;

define: sgsf93=sgsf93+3000000*gender;

analysis:
type = twolevel mixture;
process = 8;
estimator = ml;

model:
%within%
%overall%
gw by u1@1
u2-u13* (lam2-lam13);
%c#1%
gw*1;
%c#2%
gw*1;
%between%
%overall%
gb by u1@1
u2-u13* (lam2-lam13);
%c#1%
[gb@0]; gb*1;
%c#2%
[gb*0]; gb*1;
```

26 Appendix V: TOCA Model H4.

```
variable:
names = ....
usevariables = u1-u13 gender gender2;
categorical = u1-u13;
cluster = sgsf93;
within= u1-u13 gender gender2;

define:gender2=1-gender;

analysis:
type = twolevel random;
process = 8;
estimator = ml;

model:
%within%
gw by u1-u13;
gw*1; [gw@0];
s1 | gw on gender;
s2 | gw on gender2;
%between%
[s1@0];
[s2*1];
s1*1;
s2*1;
s1 with s2*0;
```

27 Appendix W: TOCA Model H5.

```
variable:
names = ....
usevariables = u1-u13 gender cl2 one;
categorical = u1-u13;
cluster = sgsf93 cl2;
within= gender u1-u13 one;

data: file= ....; variance=nocheck;

define: cl2=sgsf93+3000000*gender; one=1;

analysis:
type = threelevel random;
process = 2; thin=10;
estimator = bayes;
biter=(10000);

model:
%within%
gw by u1@1
u2-u13* (lam2-lam13);
gw*1; [gw@0];
gw on gender;
s | gw on one;
%between cl2%
s*1;
%between sgsf93%
s*1; [s@0];
```

28 Appendix X: TOCA Model H5a.

```
variable:
names = ....
usevariables = u1-u13 gender cl2;
categorical = u1-u13;
cluster = sgsf93 cl2;
within= gender;

define: cl2=sgsf93+3000000*gender;

analysis:
type = threelevel;
process = 2; thin=10;
estimator = bayes;
biter=(10000);

model:
%within%
gw by u1@1
u2-u13* (lam2-lam13);
gw*1; [gw@0];
gw on gender;
%between cl2%
gw2 by u1@1
u2-u13* (lam2-lam13);
gw2*1;
%between sgsf93%
gw3 by u1@1
u2-u13* (lam2-lam13);
gw3*1;
```

29 Appendix Y: TOCA Model H6.

```
variable:
names = ....
usevariables = u1-u13;
categorical = u1-u13;
cluster = sgsf93;
classes = c(2);
knownclass = c (gender = 0 1);

analysis:
type = twolevel mixture;
process = 2;
estimator = ml;

model:
%within%
%overall%
gw by u1@1
u2-u13* (lam2-lam13);
%c#1%
gw*1;
%c#2%
gw*1;
%between%
%overall%
gb by u1@1
u2-u13* (lam2-lam13);
gb*1;
%c#1%
[gb@0];
%c#2%
[gb*0];
```

30 Appendix Z: TOCA Model H7.

```
variable:
names = ....
usevariables = u1-u13 gender;
within=gender;
categorical = u1-u13;
cluster = sgsf93;

analysis:
type = twolevel;
process = 8;
estimator = ml;

model:
%within%
gw by u1@1
u2-u13* (lam2-lam13);
gw*1;
gw on gender;
%between%
gb by u1@1
u2-u13* (lam2-lam13);
gb*1;
```