

Modeling strategy summary

- 1) Focus on the research questions and hypotheses. It is self-evident that all effects stated in the research questions and the hypotheses should be in the model. On top of that, there may be effects that should be controlled for.
- 2) In a multilevel model one can distinguish 'explaining' the scores of a dependent variable at the first level, and 'explaining' variation between intercepts and slopes at the second level. Both level-1 and level-2 predictors can be used to explain intercept variance. Cross-level interactions (interactions between level-2 and level-1 predictors) serve to explain slope variance. Level-2 predictors, of course, cannot explain level-1 variance.
- 3) In general, when a model has an interaction effect, make sure that the corresponding 'main effects' are also in the model.
- 4) Allow a predictor always a fixed effect. In addition, level-1 predictors may be given a random effect at the second level (a random slope).
- 5) For a predictor at level 1, it may be important to question whether the effect of the predictor at the individual level within groups (within level-2 units) is the same as the effect of the same predictor at the group level. This can be checked by entering at level 2 a variable holding the group means of this predictor. If these group means have a significant effect on the dependent variable, then it is clear that the between-groups regression differs from the within-groups regression.
- 6) When two or more random effects are present in the model, then one should consider whether or not to include the covariance(s) between these random effects as a free parameter in the model. MLwiN automatically includes all covariances when random slopes are defined. Maybe, some of these should be deleted from the model. Always keep the covariance between a random slope and the intercept in the model. When there are two or more random slopes, then there may be reasons to leave out (some) slope by slope covariances. In case of random slopes for a set of dummy variables that together represent one categorical variable, one should keep all covariances between random slopes of these dummies in the model.
- 7) It often makes sense to simplify a model by removing non-significant effects. But there may be reasons to keep a non-significant effect in the model. These reasons may be related for instance with issue 1), issue 3), and issue 6). A set of dummy variables representing one categorical variable is in general best treated as a conceptual unity; so don't remove non-significant dummies from a set that includes a dummy variable with a significant coefficient, or test the set as a whole for significance using the deviance test.
- 8) Multicollinearity (too high correlations between effects in a model) can be a big problem in multilevel analysis, especially when the number of predictors is relatively high compared with N. This can easily happen at a higher level.
- 9) Cross-level interactions are meant to explain variation in slopes. This does not necessarily imply that you should always first show significant slope variance before allowing a cross-level interaction in the model. There might be theoretical considerations to expect or to test for a specific cross-level interaction.
- 10) Test results concerning a certain effect may depend on other effects in the model; one should be aware of that when deciding to include or exclude an effect.
- 11) With every statistical test one can make a Type-I error or a Type-II error. Low power can be especially bothersome at higher levels.