

**Assignment 1**  
**First draft due October 16, 2007**

For this assignment, you will utilize data from the National Educational Longitudinal Study (1988). The dataset and codebook can be found at <http://www.sarkisian.net/sc705/datasets.html>. There, you will find two zipped folders with SPSS and Stata datasets, as well as a folder containing smaller versions of the data (to be used with the student version of HLM 6) in both SPSS and Stata formats. These datasets contain information on students nested within schools (as a single data file). You will also find a text file containing the codebook.

- 1) Select a continuous dependent variable and 4-5 potential independent variables on each level. Note that at least one of your level-2 independent variables has to be an aggregate of a level-1 variable that will also be included in your analysis; if desired, other variables can be aggregates as well. Devise a theoretical argument you can test using these data.
- 2) Start a running log that will contain the output (with your brief comments) for all of the tasks that you will perform for this assignment. You will have to assemble that log file manually (in your Word processor) as you will go between your main statistical software (e.g. SPSS or Stata) and HLM 6, paste various pieces of syntax and output, and provide brief comments. Those of you who work with SPSS might be unaccustomed to documenting what you do while using the menus; make sure you paste your commands into a syntax file before executing them!
- 3) Data management: Using either SPSS or Stata data file, keep ID variables and your variables of interest in the dataset. Drop the rest. Check for missing data; if you have missing data on your variables of interest for certain schools, drop those schools. Generate aggregate level-2 variables (using level-1 variables) that you will need for your analyses. When finished with data management, import the data into HLM (create an MDM file). Pay attention to level 1 and level 2 missing data, but do not attempt any imputation. Examine statistics to make sure that importing went correctly (include these statistics and the corresponding statistics from Stata or SPSS in your log).
- 4) Examine the data graphically in HLM.
- 5) For the variables specified by your theoretical argument, build the main types of models that we discussed, gradually increasing the model complexity. The last model you estimate in this step should be a random coefficients (intercepts and slopes) model with level 2 predictors (some of them aggregates of level 1 predictors) used to predict both the intercepts and slopes. For all models, make appropriate centering choices.
- 6) Based on the results of the models built in #5, trim the random coefficients model. Utilize hypothesis testing when trimming the model.
- 7) For your final model, perform the range of diagnostics addressed in class (most of them you will have to do using another statistical package, e.g. SPSS or Stata). If necessary, consider remedies.

8) Interpret the results of your final model, both the fixed effects and the random components. Use graphs to assist your interpretation.

9) Write up the results of your final model in journal format. First, include an Introduction that will provide a short substantive description of your theoretical argument, your research questions, and hypotheses (1 page max.). Second, include a brief Data and Methods section (1-2 pages) describing the variables and the methodology. Include any discussion of diagnostics and modifications in this section. Also, include a table with descriptive statistics for your sample. Third, provide a 3-4 page description of the results including tables (in journal format) and graphs assisting in the interpretation of results (do not show all tables and graphs – select those most helpful for understanding the results). If it is absolutely necessary that you discuss some diagnostics or model modifications here rather than in the methods section, do so using footnotes. The page limit for this assignment is 7 pages double-spaced.

10) Along with your assignment, please include an appendix containing your running log – i.e., your data management efforts and the output for all of your models and diagnostics, with brief comments. Make sure that your output includes all the steps specified in items 1-8 in this assignment. There is no page limit for your output but please edit it to contain only the relevant syntax, output, and graphs (i.e., omit any unproductive steps).