



## Three-level model for the EG data

1. Constructing the MDM file from raw data .....	1
1.1 SPSS input.....	1
1.2 ASCII input.....	8
1.3 Other file input .....	8
2. Creating the command file.....	8
3. Interpreting the results .....	10

### 1. Constructing the MDM file from raw data

In constructing the MDM file, the user has the same range of options for data input for HLM3 as for HLM2. We first describe the use of SPSS file input and then consider ASCII, SYSTAT, SAS, and other data file formats.

#### 1.1 SPSS input

Data input requires a level-1 file (in our illustration a time-series data file), a level-2 file (child-level file), and a level-3 (school-level) file.

##### Level-1 file

The level-1 file, EG1.SAV, has 7242 observations collected on 1721 children beginning at the end of grade one and followed up annually thereafter until grade six. There are four level-1 variables (not including the schoolid and the childid). Time-series data for the first two children are shown in below.

There are eight records listed, three for the first child and five for the second. (Typically there are four or five observations per child with a maximum of six.) The first ID is the level-3 (*i.e.*, school) ID and the second ID is the level-2 (*i.e.*, child) ID. We see that the first record comes from school 2020 and child 273026452 within that school. Notice that this child has three records, one for each of three measurement occasions. Following the two ID fields are that child's values on four variables:

- YEAR (year of the study minus 3.5)  
This variable can take on values of  $-2.5$ ,  $-1.5$ ,  $-0.5$ ,  $0.5$ ,  $1.5$ , and  $2.5$  for the six years of data collection.
- GRADE  
The grade level minus 1.0 of the child at each testing occasion. Therefore, it is 0 at grade 1, 1 at grade 2, etc.
- MATH

A math test in an IRT scale score metric.

- RETAINED

An indicator that a child is retained in grade for a particular year (1 = retained, 0 = not retained).

	schoolid	childid	year	grade	math	retained
1	2020	273026452	.50	2.00	1.15	.00
2	2020	273026452	1.50	3.00	1.13	.00
3	2020	273026452	2.50	4.00	2.30	.00
4	2020	273030991	.50	2.00	2.43	.00
5	2020	273030991	1.50	3.00	2.25	.00
6	2020	273030991	2.50	4.00	3.87	.00
7	2020	273030991	-.50	1.00	.44	.00
8	2020	273030991	-1.50	.00	-1.30	.00

We see that the first child, child 27306452 in school 2020, had values of 0.5, 1.5, and 2.5 on year. Clearly, that child had no data at the first three data collection waves (because we see no values of  $-2.5$ ,  $-1.5$ , or  $-0.5$  on year), but did have data at the last three waves. We see also that this child was not retained in grade during this period since the values for GRADE increase by 1 each year and since RETAINED takes on a value of 0 for each year. The three MATH scores of that child (1.15, 1.13, 2.30) show no growth in time period 1.5. Oddly enough, the time-series record for the second child (child 273030991 in school 2020) displays a similar pattern in the same testing.

**Note: The level-1 and level-2 files must also be sorted in the same order of level-2 ID nested within level-3 ID, e.g., children within schools. If this nested sorting is not performed, an incorrect multivariate data matrix file will result.**

### Level-2 file

The level-2 units in the illustration are 1721 children. The data are stored in the file EG2.SAV. The level-2 data for the first eight children are listed below. The first field is the schoolid and the second is the childid. Note that each of the first ten children is in school 2020.

There are three variables:

- FEMALE (1 = female, 0 = male)
- BLACK (1 = African-American, 0 = other)
- HISPANIC (1= Hispanic, 0 = other)

We see, for example, that child 273026452 is a Hispanic male (FEMALE = 0, BLACK = 0, HISPANIC = 1).

	<b>schoolid</b>	<b>childid</b>	<b>female</b>	<b>black</b>	<b>hispanic</b>
1	2020	273026452	.00	.00	1.00
2	2020	273030991	.00	.00	.00
3	2020	273059461	.00	.00	1.00
4	2020	278058841	.00	.00	.00
5	2020	292017571	.00	.00	1.00
6	2020	292020281	.00	.00	.00
7	2020	292020361	.00	.00	.00
8	2020	292025081	.00	.00	.00

### Level-3 file

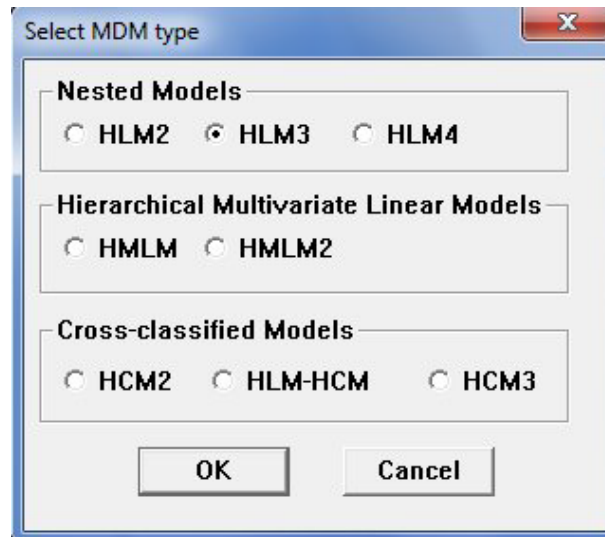
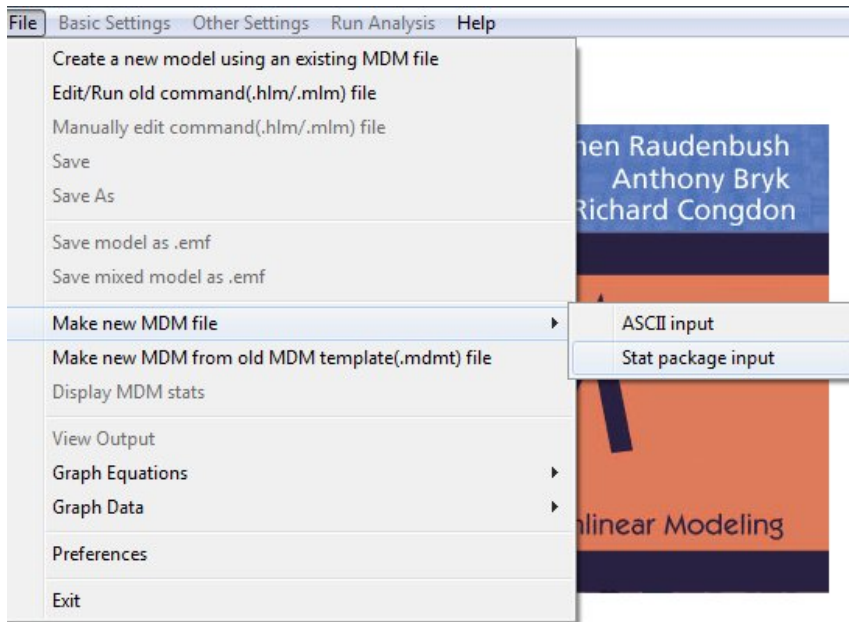
The level-3 units in the illustration are 60 schools. Level-3 data for the first seven schools are printed below. The full data are in the file EG3.SAV. The first field on the left is the schoolid. There are three level-3 variables:

- SIZE, number of students enrolled in the school
- LOWINC, the percent of students from low income families
- MOBILE, the percent of students moving during the course of a single academic year

We see that the first school, school 2020, has 380 students, 40.3% of whom are low income. The school mobility rate is 12.5%.

	<b>schoolid</b>	<b>size</b>	<b>lowinc</b>	<b>mobility</b>
1	2020	380.00	40.30	12.50
2	2040	502.00	83.10	18.60
3	2180	777.00	96.60	44.40
4	2330	800.00	78.90	31.70
5	2340	1133.00	93.70	67.00
6	2380	439.00	36.90	39.30
7	2390	566.00	100.00	39.90

In sum, there are four variables at level 1, three at level 2 and three at level 3. Note that the ID variables do not count as variables. Note that the three files are linked by level-2 and level-3 IDs here. The user first informs HLM that the input files are SPSS system files and the MDM is a three-level file.



Then HLM is supplied with the appropriate information for the data. The user must identify the two sets of IDs, the number of variables in each file, the variable names, and the filenames. Start by clicking Browse to look for the data file. Note that the **Input File Type** is by default set to **SPSS/Windows**, which happens to be the format we are using. For other types of data, the correct option should be selected from this drop-down menu before browsing for the data.

**Make MDM HLM3**

MDM template file

File Name:

MDM File Name (use .mdm suffix)

Input File Type

Structure of Data - this affects the notation only!

cross sectional  cross-sectional with measurement model at level-1

longitudinal  longitudinal with measurement model at level-1

Level-1 Specification

Level-1 File Name:

Missing Data?  No  Yes Delete missing level-1 data when:  making mdm  running analyses

Level-2 Specification

Level-2 File Name:

Level-3 Specification

Level-3 File Name:

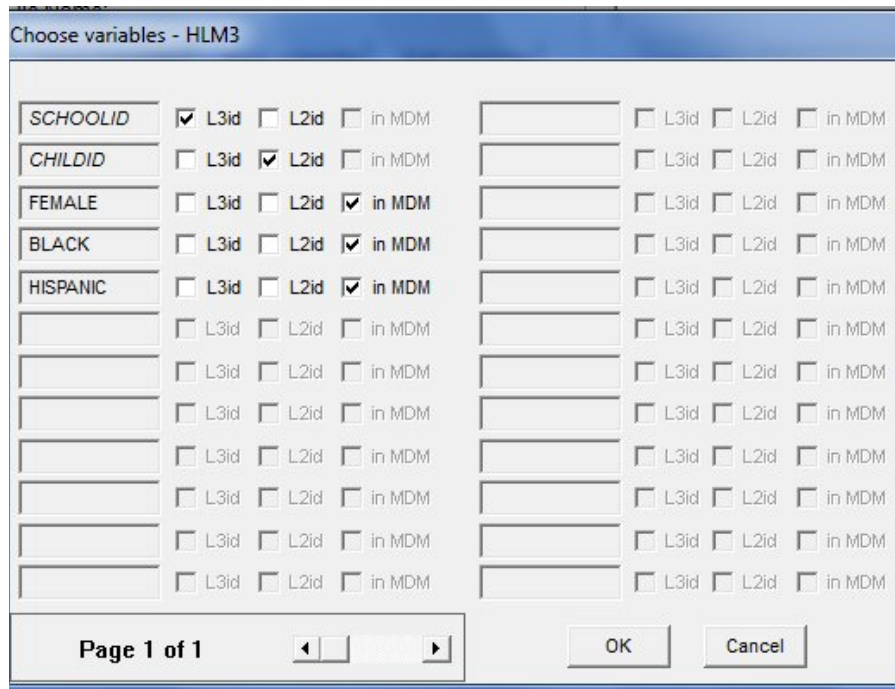
Once the data file has been selected, the **Choose variables** dialog box is opened. In the case of the level-2 and level-1 data files (if separate files are used) the level-3 and level-2 ID should be identified first by checking the appropriate box. Other variables of interest are selected in the in MDM column as shown below for the level-1 data file **EG1.SAV**. Note that only the ID variables may be in string/character format: all other variables must be in numeric format.

**Choose variables - HLM3**

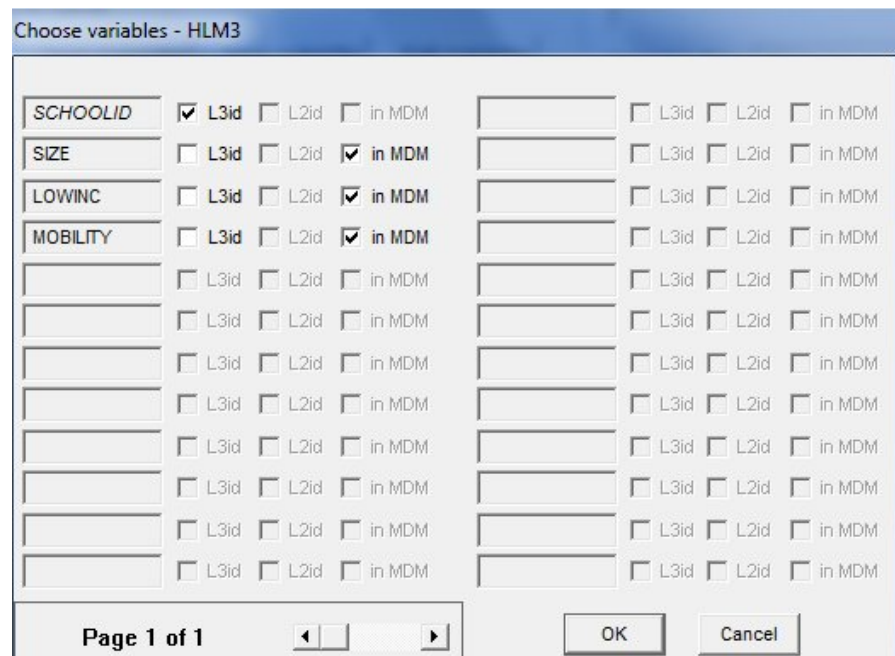
SCHOOLID	<input checked="" type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
CHILDID	<input type="checkbox"/> L3id	<input checked="" type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
YEAR	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input checked="" type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
GRADE	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input checked="" type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
MATH	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input checked="" type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
RETAINED	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input checked="" type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM
	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM	<input type="text"/>	<input type="checkbox"/> L3id	<input type="checkbox"/> L2id	<input type="checkbox"/> in MDM

Page 1 of 1

Selection of variables at level-2 are done in a similar way.

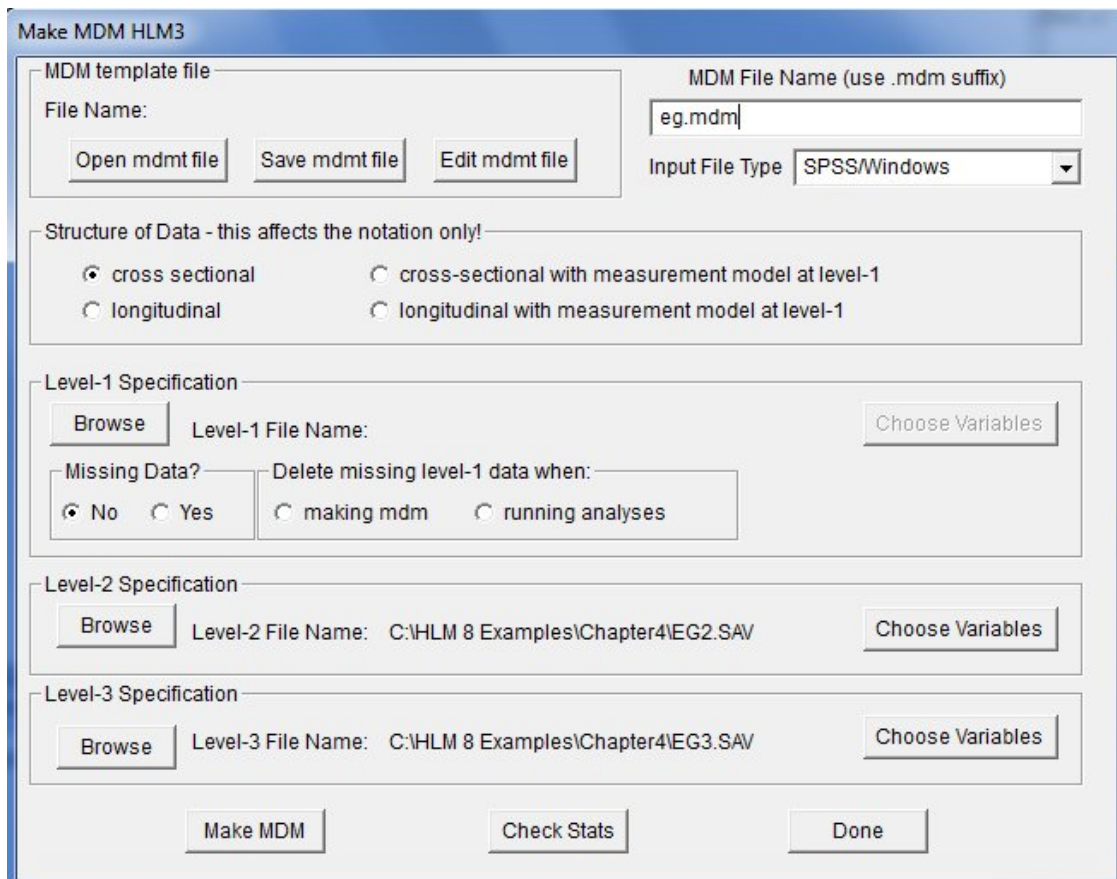


The level-3 data file only has to contain the level-3 ID – a level-2 ID is not necessary. At any level, at least one variable other than the ID must be selected.



If there are missing data in the file, the **Missing Data** field must be updated to **Yes**, and the time of deletion (**making MDM or running analysis**) specified. The program can handle **missing data at level-1 only**. HLM3 will listwise delete cases with missing data at levels two and three. Before saving the MDMT file, provide a name for the MDM file. The completed dialog box for the EG data is shown below.

The response file, EGSPSS.MDMT, contains a log of the input responses used to create the MDM file, EG.MDM, using EG1.SAV, EG2.SAV, and EG3.SAV.



Next, click the **Make MDM** button to create the MDM file. If the file is created successfully, the descriptive statistics of the contents of the file will automatically be displayed in Notepad. These statistics can also be accessed by clicking the **Check Stats** button at the bottom of the **Make MDM** dialog box.

LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
YEAR	7230	0.38	1.39	-2.50	2.50
GRADE	7230	1.81	1.35	0.00	5.00
MATH	7230	-0.54	1.53	-5.22	5.77
RETAINED	7230	0.05	0.22	0.00	1.00

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
FEMALE	1721	0.51	0.50	0.00	1.00
BLACK	1721	0.69	0.46	0.00	1.00
HISPANIC	1721	0.15	0.35	0.00	1.00

LEVEL-3 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SIZE	60	642.53	317.37	113.00	1486.00
LOWINC	60	73.74	27.27	0.00	100.00
MOBILITY	60	34.75	13.21	8.80	67.00

MDM template: C:\HLM 8 Examples\Chapter4\EG.MDMT  
MDM file name: EG.MDM

After constructing the MDM file, you should check whether the data have been properly read into HLM by examining the descriptive statistics of the MDM file.

## 1.2 ASCII input

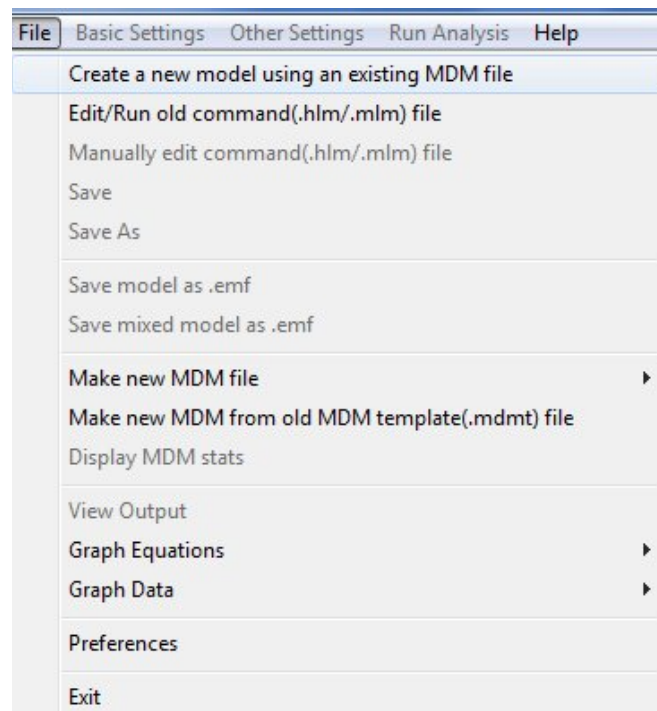
The procedure for constructing an MDM file from ASCII data files is similar to that for SPSS file input. The major difference is that the format statements must be entered for the three data files, variable names, and missing value codes, if applicable. Rules about the format are included in the Appendix of the manual distributed with the program.

## 1.3 Other file input

For SAS and SYSTAT file input, a user selects either SAS5 transport or SYSTAT from the **Input File Type** drop-down list box as appropriate before clicking the **Browse** buttons in the file specification sections and follows the same steps for SPSS input type to create MDM files. HLM3 has the same range of options for data input as HLM2. In addition to SYSTAT, SPSS, STATA, free format, and SAS, the Windows version (through a third-party module) allows numerous other data formats from, for example, EXCEL, and LOTUS input.

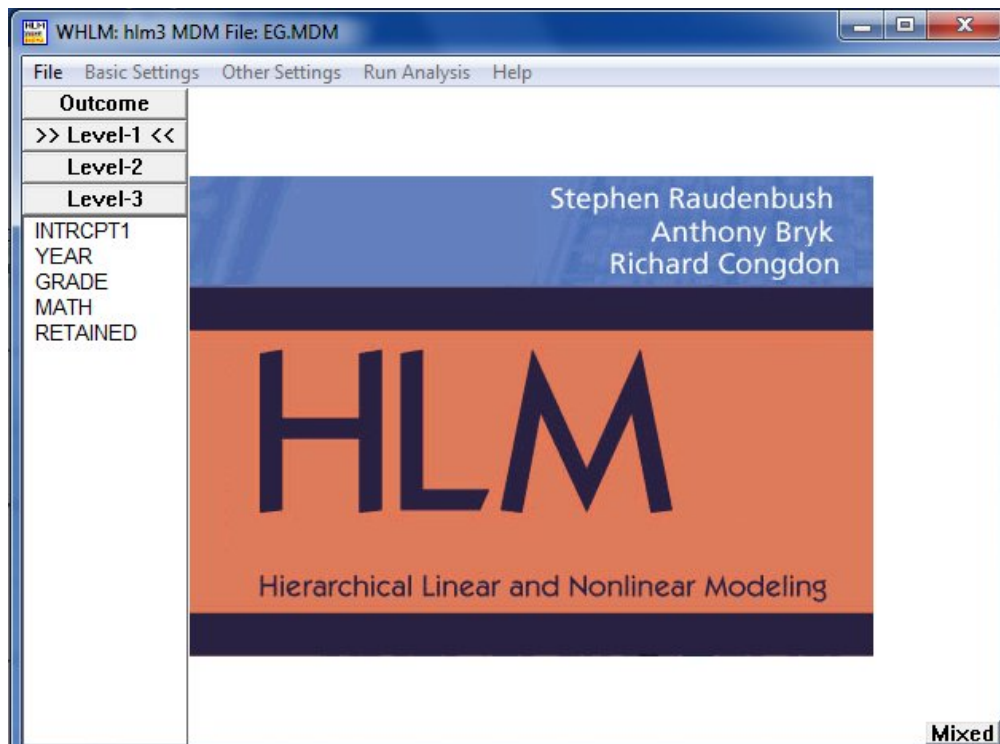
## 2. Creating the command file

Start by selecting the **Create a new model using an existing MDM file** option from the **File** menu.

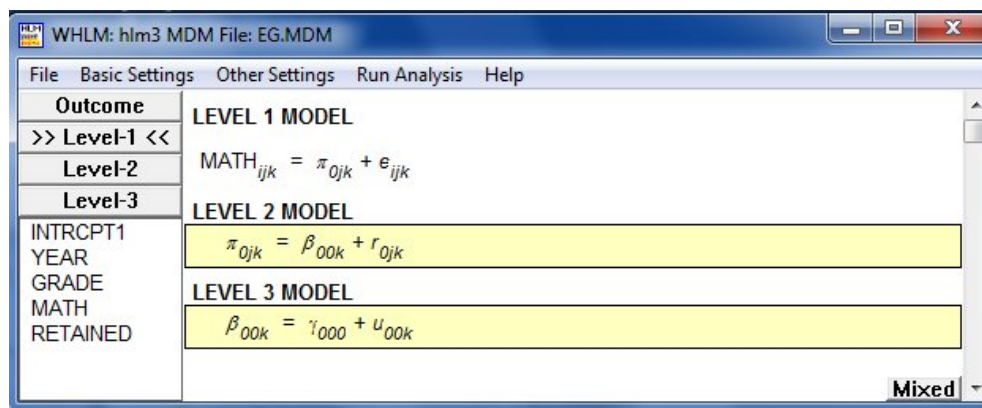


Browse for the MDM file **EG.MDM** created in the previous section and select it to return to the main window, where the variables included in the MDM file are now displayed at the left of the main window.

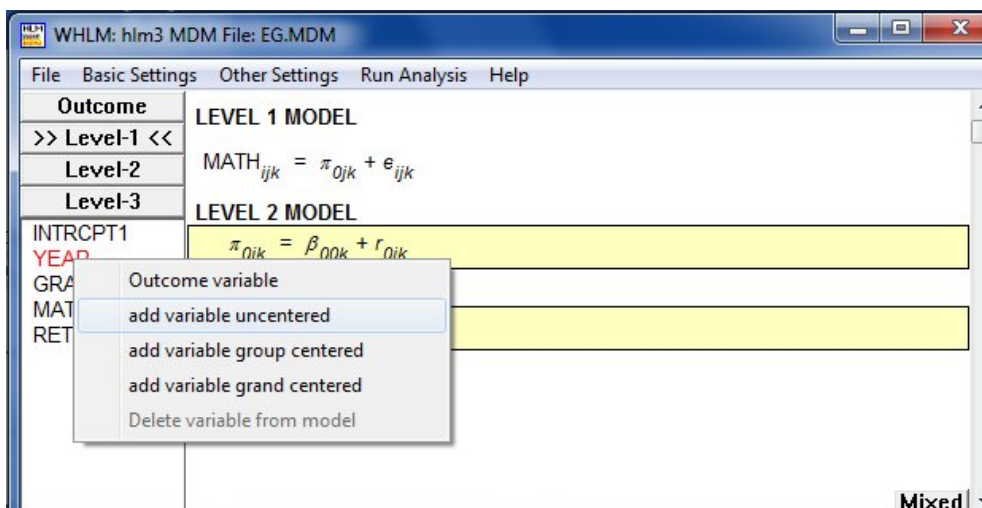




Start by selecting the variable MATH, representing the student's mathematics score, as outcome variable.



We wish to explore the relationship between mathematics achievement and time of measurement. Click on the variable name YEAR and select it as level-1 predictor of the mathematics achievement MATH.



The completed model is displayed below.

In this model, each level-1 coefficient – the intercept and YEAR slope- becomes an outcome variable. For each of these, child characteristics may be used as level-2 predictors. In principle, the level-2 parameters then describe the distribution of growth curves within each school.

In turn, each level-2 coefficient becomes an outcome variable at level-3. Here school characteristics may be included as predictors on any of the equations to predict school-to-school variation in these level-2 coefficients. The level-3 model specifies how schools differ with respect to the distribution of growth curves within them.

By default, only the intercept will be assumed to be varying randomly over the higher-level units. Activate the random slope effects at level-2 and level-3 by clicking on  $r_{1jk}$  and  $u_{10k}$  respectively.

At the bottom of the main window, the model is displayed in mixed model formulation.

WHLM: hlm3 MDM File: EG.MDM Command File: newcmd.hlm

File Basic Settings Other Settings Run Analysis Help

**Outcome** LEVEL 1 MODEL

**Level-1**

>> **Level-2** <<

**Level-3**

INTRCPT2  
FEMALE  
BLACK  
HISPANIC

LEVEL 2 MODEL

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$

LEVEL 3 MODEL

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{10k} = \gamma_{100} + u_{10k}$$

Mixed

Mixed Model

$$\text{MATH}_{ijk} = \gamma_{000} + \gamma_{100} * \text{YEAR}_{ijk} + r_{0jk} + r_{1jk} * \text{YEAR}_{ijk} + u_{00k} + u_{10k} * \text{YEAR}_{ijk} + e_{ijk}$$

Save the command file using the **File, Save As** option from the main menu bar before running the analysis. Once the iterative procedure has converged, output will automatically be displayed in the user's viewer of choice (set on the **File, Preferences** menu).

### 3. Interpreting the results

In the first section of the output file, Least Squares results are given. These are used to provide starting values for parameters to the program and should not be used to base conclusions on.

#### Least Squares Estimates

$$\sigma^2 = 1.21432$$

**Least-squares estimates of fixed effects**

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, $\pi_0$					
For INTRCPT2, $\beta_{00}$					
INTRCPT3, $\gamma_{000}$	-0.827685	0.013431	-61.623	7228	<0.001
For YEAR slope, $\pi_1$					
For INTRCPT2, $\beta_{10}$					
INTRCPT3, $\gamma_{100}$	0.765828	0.009293	82.410	7228	<0.001

**Least-squares estimates of fixed effects (with robust standard errors)**

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, $\pi_0$					
For INTRCPT2, $\beta_{00}$					
INTRCPT3, $\gamma_{000}$	-0.827685	0.072631	-11.396	7228	<0.001
For YEAR slope, $\pi_1$					
For INTRCPT2, $\beta_{10}$					
INTRCPT3, $\gamma_{100}$	0.765828	0.018892	40.537	7228	<0.001

The least-squares likelihood value = -1.096090E+04

Deviance = 21921.80879

Number of estimated parameters = 3

For starting values, data from 7230 level-1 and 1721 level-2 records were used

**Starting Values**

$\sigma^2_{(0)} = 0.29710$

$T_{\pi(0)}$

INTRCPT1,  $\pi_0$  0.71125 0.05143

YEAR,  $\pi_1$  0.05143 0.01582

$T_{\beta(0)}$

INTRCPT1 YEAR

INTRCPT2,  $\beta_{00}$  INTRCPT2,  $\beta_{10}$

0.14930 0.01473

0.01473 0.01196

Final results after convergence are given next. The first results give estimates of the variance and covariance components. From the elements of the  $\tau_{\pi}$  (as correlations) matrix, we note that the estimated correlation between true status at YEAR = 3.5 (halfway through 3<sup>rd</sup> grade) and true rate of change is estimated to be 0.551 for children in the same school.

Final Results - Iteration 9

**Iterations stopped due to small change in likelihood function**

$\sigma^2 = 0.30148$

Standard error of  $\sigma^2 = 0.00660$

$\tau_{\pi}$		
INTRCPT1, $\pi_0$	0.64049	0.04676
YEAR, $\pi_1$	0.04676	0.01122
Standard errors of $\tau_{\pi}$		
INTRCPT1, $\pi_0$	0.02515	0.00499
YEAR, $\pi_1$	0.00499	0.00196
$\tau_{\pi}$ (as correlations)		
INTRCPT1, $\pi_0$	1.000	0.551
YEAR, $\pi_1$	0.551	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, $\pi_0$	0.839
YEAR, $\pi_1$	0.190

$\tau_{\beta}$			
INTRCPT1	YEAR		
INTRCPT2, $\beta_{00}$	INTRCPT2, $\beta_{10}$		
0.16531	0.01705		
0.01705	0.01102		
Standard errors of $\tau_{\beta}$			
INTRCPT1	YEAR		
INTRCPT2, $\beta_{00}$	INTRCPT2, $\beta_{10}$		
0.03641	0.00720		
0.00720	0.00252		
$\tau_{\beta}$ (as correlations)			
INTRCPT1/INTRCPT2, $\beta_{00}$	1.000	0.399	
YEAR/INTRCPT2, $\beta_{10}$	0.399	1.000	

Random level-2 coefficient	Reliability estimate
INTRCPT1/INTRCPT2, $\beta_{00}$	0.821
YEAR/INTRCPT2, $\beta_{10}$	0.786

Inspection of the  $\tau_{\beta}$  (as correlations) matrix lead to the conclusion that the estimated correlation between true school mean status at YEAR = 3.5 and true school-mean rate of change is 0.399.

Reliabilities of school-level parameter estimates are also given. These indicate the reliability with which we can discriminate among level-2 units using their least-squares estimates of  $\beta_0$  and  $\beta_1$ . Low reliabilities do not invalidate the HLM analysis. Very low reliabilities (e.g., < 0.10), often indicate that a random coefficient might be considered fixed in subsequent analyses.

This is followed by the final results for the fixed effects.

**Final estimation of fixed effects:**

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, $\pi_0$					
For INTRCPT2, $\beta_{00}$					
INTRCPT3, $\gamma_{000}$	-0.779309	0.057829	-13.476	59	<0.001
For YEAR slope, $\pi_1$					
For INTRCPT2, $\beta_{10}$					
INTRCPT3, $\gamma_{100}$	0.763029	0.015263	49.993	59	<0.001

**Final estimation of fixed effects (with robust standard errors)**

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, $\pi_0$					
For INTRCPT2, $\beta_{00}$					
INTRCPT3, $\gamma_{000}$	-0.779309	0.057830	-13.476	59	<0.001
For YEAR slope, $\pi_1$					
For INTRCPT2, $\beta_{10}$					
INTRCPT3, $\gamma_{100}$	0.763029	0.015260	50.000	59	<0.001

The first table provides model-based estimates of the standard errors while the second table provides robust estimates of the standard errors. Note that the two sets of standard errors are similar. If the robust and model-based standard errors are substantively different, it is recommended that the tenability of key assumptions should be investigated further. Robust errors cannot be produced if sufficient data are not available.

The estimate of 0.763 for  $\gamma_{100}$  indicates a significant positive average growth rate at 0.763 logits per year.

**Final estimation of level-1 and level-2 variance components**

Random Effect	Standard Deviation	Variance Component	d.f.	$\chi^2$	p-value
INTRCPT1, $r_0$	0.80030	0.64049	1661	13679.62589	<0.001
YEAR slope, $r_1$	0.10595	0.01122	1661	2132.50756	<0.001
level-1, e	0.54907	0.30148			

**Final estimation of level-3 variance components**

Random Effect	Standard Deviation	Variance Component	d.f.	$\chi^2$	p-value
INTRCPT1/INTRCPT2, $u_{00}$	0.40658	0.16531	59	488.30922	<0.001
YEAR/INTRCPT2, $u_{10}$	0.10498	0.01102	59	377.43020	<0.001

**Statistics for the current model**

Deviance = 16326.231407

Number of estimated parameters = 9

The final set of results in the output file concern the random effects included in the model. We note that there is significant variability in math achievement scores in terms of mean status at YEAR = 3.5 with  $\hat{\text{var}}(u_{00}) = 0.16531$ .

The same is true in terms of school-mean rates of change with  $\hat{\text{var}}(u_{10}) = 0.01102$ . The same holds true at the student level (see the estimates of  $\hat{\text{var}}(r_0)$  and  $\hat{\text{var}}(r_1)$  in the first table).