



Three-level model with exploratory analysis at level-3 for the EG data

1. The data.....	1
2. Creating the command file.....	3
3. Interpreting the results.....	7

1. The data

This is the second of four HLM3 examples based on the EG data. A brief description of the data contained in the MDM file **EG.MDM** is given below. The command file **EG1.MLM** created in the first example is used here.

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).

	schoolid	childid	female	black	hispanic
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%.

	schoolid	size	lowinc	mobility
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

2. Creating the command file

Start by selecting the **Edit/Run old command (.hlm/.mlm) file** option from the **File** menu. Browse for **EG1.MLM** and click **OK** to return to the main window where this model is now displayed.

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

File Basic Settings Other Settings Run Analysis Help

Outcome

Level-1

>> **Level-2** <<

Level-3

INTRCPT2
FEMALE
BLACK
HISPANIC

LEVEL 1 MODEL

$$\text{MATH}_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{YEAR}_{ijk}) + \epsilon_{ijk}$$

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

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

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.

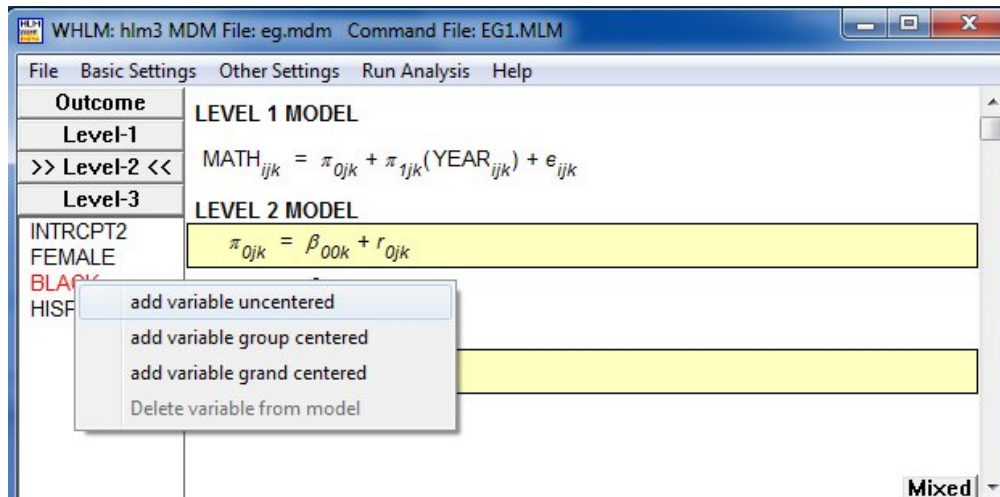
In turn, each level-2 coefficient becomes an outcome variable at level-3. Here school characteristics may be included as predictors on any of the equations to predict school-to-school variation in these level-2 coefficients. The level-3 model specifies how schools differ with respect to the distribution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

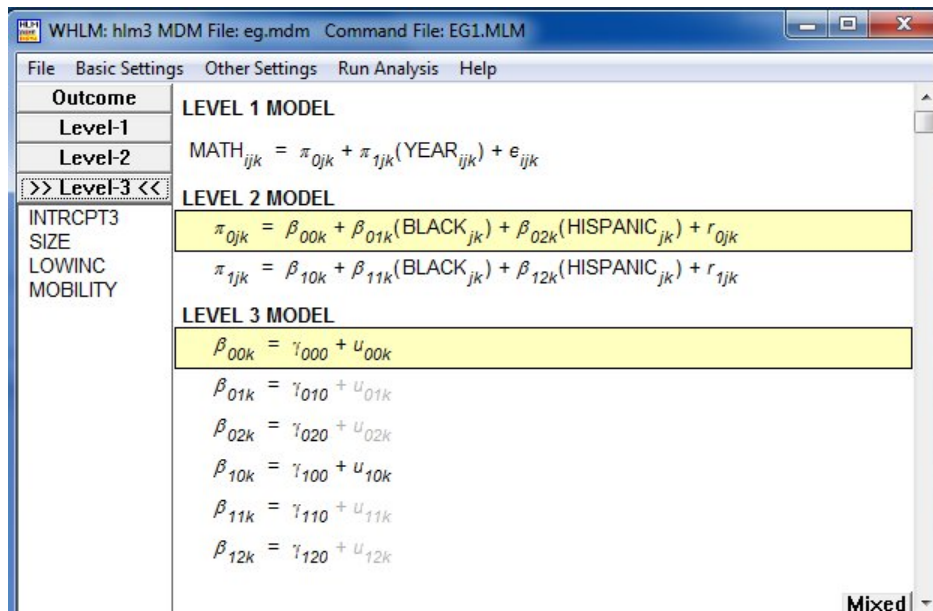
At the bottom of the main window, the model is displayed in mixed model formulation.

We wish to extend this model by including student level and school level predictors, simultaneously also evaluating the “t-to-enter statistics” of additional predictors.

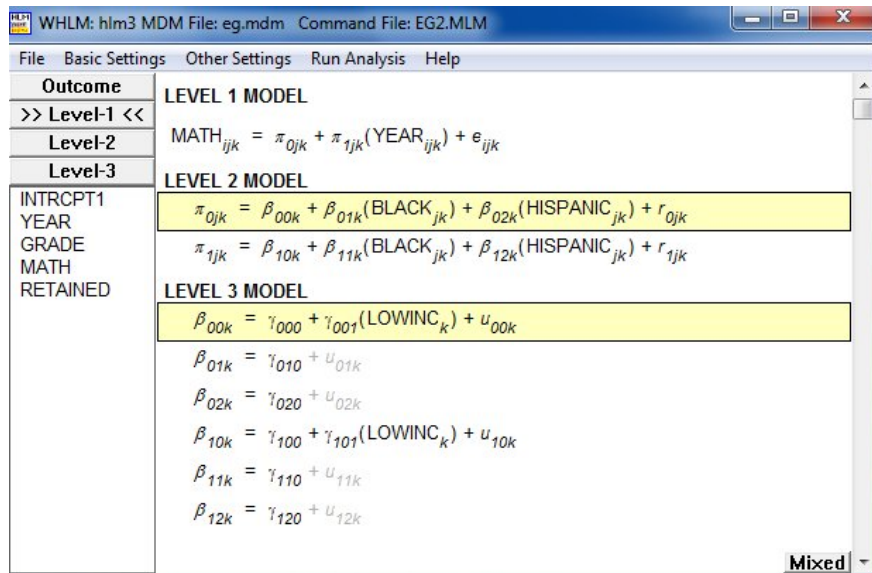
The first step is to complete the model. Click on the >>**Level-2**<< button on the left of the main window to display available level-2 predictors. Add BLACK and HISPANIC to the active level-2 equation (for the intercept) by selecting the add variable uncentered option from the pop-up menu. Click on the second level-2 equation, associated with the YEAR slope, and add both predictors to this equation as well.



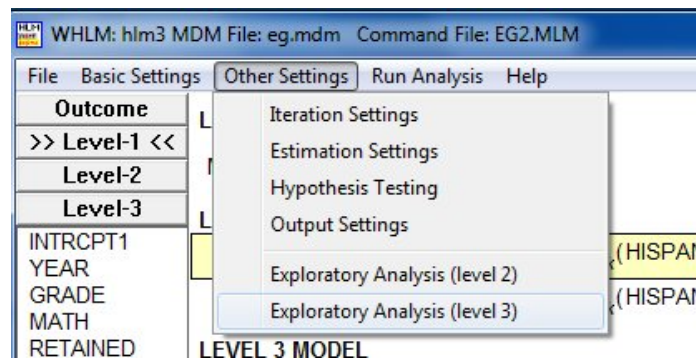
This completes specification of the level-2 model. Now click on the >>Level-3<< button to display the names of variables available for inclusion at this level.



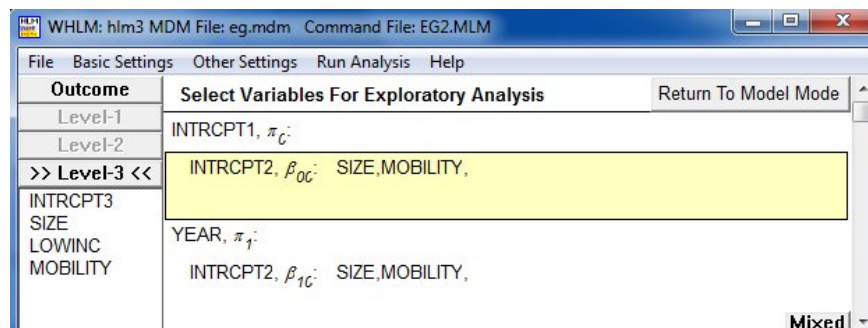
Add the level-3 variable LOWINC as predictor on the intercept equation (β_{00k}), then click on the equation for β_{10k} and add LOWINC as predictor to the YEAR slope equation as well. The completed model specification is shown below.



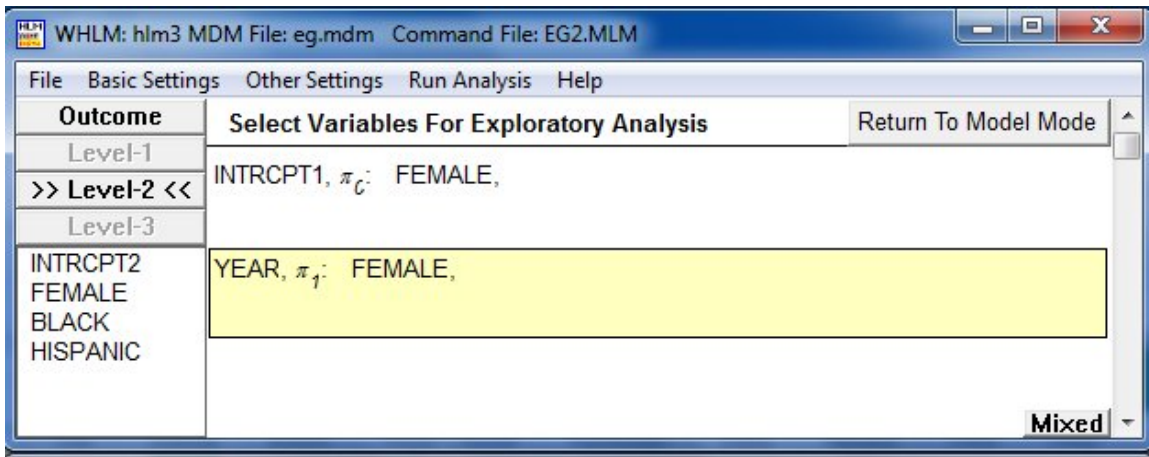
In addition to estimating this more complex model utilizing student and school level predictors, we also wish to evaluate the potential of some of the remaining level-2 and level-3 predictors. To obtain “t-to-enter” statistics for potential level-3 predictors, select the **Exploratory Analysis (level 3)** option from the **Other Settings** menu on the main men bar.



In the **Select Variables for Exploratory Analysis** window, click on the variables SIZE and MOBILITY and use the add option on the pop-up menu to add them to the active equation. Click on the second equation and repeat these actions before clicking **Return to Model Mode** and returning to the main window.



We now do the same at level-2 by selecting the **Exploratory Analysis (level 2)** option instead. At level-2, the variable FEMALE is of interest and is entered on both intercept and slope equation to evaluate it’s potential as predictor of both intercept and YEAR slope.



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.

Final results after convergence are given next. The first results give estimates of the variance and covariance components. From the elements of the τ_{π} (as correlations) matrix, we note that the estimated correlation between true status at YEAR = 3.5 (halfway through 3rd grade) and true rate of change is estimated to be 0.558 for children in the same school.

Final Results - Iteration 64 Iterations stopped due to small change in likelihood function

$$\sigma^2 = 0.30141$$

T_{π}
 INTRCPT1, π_0 0.62233 0.04657
 YEAR, π_1 0.04657 0.01121

T_{π} (as correlations)
 INTRCPT1, π_0 1.000 0.558
 YEAR, π_1 0.558 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, π_0	0.835
YEAR, π_1	0.190

T_{β}
 INTRCPT1 YEAR
 INTRCPT2, β_{00} INTRCPT2, β_{10}
 0.07809 0.00082
 0.00082 0.00798

T_{β} (as correlations)
 INTRCPT1/INTRCPT2, β_{00} 1.000 0.033
 YEAR/INTRCPT2, β_{10} 0.033 1.000

Random level-2 coefficient	Reliability estimate
INTRCPT1/INTRCPT2, β_{00}	0.702
YEAR/INTRCPT2, β_{10}	0.734

Inspection of the τ_β (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 0.033. This is considerably smaller than in the model fitted in the first example and due to the addition of predictors to this model. There is stronger correlation between true school mean status at YEAR = 3.5 and true school-mean rate of change at student level than at school level.

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

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	0.140661	0.113808	1.236	58	0.221
LOWINC, γ_{001}	-0.007578	0.001396	-5.429	58	<0.001
For BLACK, β_{01}					
INTRCPT3, γ_{010}	-0.502077	0.076837	-6.534	1597	<0.001
For HISPANIC, β_{02}					
INTRCPT3, γ_{020}	-0.319361	0.081919	-3.899	1597	<0.001
For YEAR slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	0.874528	0.037282	23.457	58	<0.001
LOWINC, γ_{101}	-0.001369	0.000499	-2.744	58	0.008
For BLACK, β_{11}					
INTRCPT3, γ_{110}	-0.030963	0.022270	-1.390	1597	0.165
For HISPANIC, β_{12}					
INTRCPT3, γ_{120}	0.043084	0.024369	1.768	1597	0.077

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

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	0.78888	0.62233	1659	13374.06588	<0.001
YEAR slope, r_1	0.10587	0.01121	1659	2128.27086	<0.001
level-1, e	0.54900	0.30141			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2, u_{00}	0.27945	0.07809	58	255.00659	<0.001
YEAR/INTRCPT2, u_{10}	0.08930	0.00798	58	276.56072	<0.001

We note that while the coefficients for the predictors BLACK and HISPANIC are statistically significant from zero on the intercept equation ($\hat{\gamma}_{010}$ and $\hat{\gamma}_{020}$) the same is not true for the YEAR slope ($\hat{\gamma}_{110}$ and $\hat{\gamma}_{120}$). At level-3, the predictor LOWINC seems to make a significant contribution to the explanation of variation in both intercept and slope ($\hat{\gamma}_{001}$ and $\hat{\gamma}_{101}$).

The next 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 $\hat{\text{var}}(u_{10}) = -.01102$. 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).

Exploratory Analysis: estimated level-2 coefficients and their standard errors obtained by regressing EB residuals on level-2 predictors selected for possible inclusion in subsequent HLM runs

Level-1 Coefficient	Potential Level-2 Predictors
INTRCPT1, π_0	
	FEMALE
Coefficient	-0.010
Standard Error	0.036
t-value	-0.274
YEAR, π_1	
	FEMALE
Coefficient	0.001
Standard Error	0.003
t-value	0.461

Exploratory Analysis: estimated level-3 coefficients and their standard errors obtained by regressing EB residuals on level-3 predictors selected for possible inclusion in subsequent HLM runs

Level-1 Coefficient	Potential Level-3 Predictors	
INTRCPT1/INTRCPT2, β_{00}		
	SIZE	MOBILITY
Coefficient	-0.000	-0.007
Standard Error	0.000	0.002
t-value	-0.717	-2.824
YEAR/INTRCPT2, β_{10}		
	SIZE	MOBILITY
Coefficient	-0.000	-0.000
Standard Error	0.000	0.001
t-value	-1.155	-0.541

The potential predictors not included in the model to be employed as significant predictors in subsequent models are indicated approximately by the “t-values” given above. Due to the metric of the school size, the actual coefficients and standard errors are too small to be printed in this instance.