Creating an MDM file and command file for unconditional 4-level model for the literacy data

1. Description of the data

To illustrate the operation of the HLM4 program, we reanalyze a subset of data from Hough, Bryk, Pinnell, Kerbow, Fountas, and Scharer (2008). Hough *et al*. used a four-level model to examine the association between school-based coaching and the development of teachers' expertise in literary instruction. The level-1 model in their study was a measurement error model associated with 1317 repeated observations on a measure of classroom instruction, which they called teaching *expertise*. (This measurement model relates the observed data to a "true" or latent score plus some error of measurement. See below.) The level-2 model represented a growth model for each teacher's "true scores" on teaching expertise, and the level-3 and level-4 models investigated the associations of the growth trajectory parameters with teacher- and school-level correlates with data from 219 teachers from 17 schools, respectively.

The example illustrates the use of a level-1 in HLM as a measurement model. In brief,

$$Y_{mtij} = \psi_{0tij} + \varepsilon_{mtij}, \quad \varepsilon_{mtij} \sim N\left(0, \sigma^2_{mtij}\right)$$

where

Y_{mtij} is the observed measure on occasion t for teacher i in school j,
ψ_{tij} is the true or latent value for teacher expertise, and
ε_{mtij} is the error of measurement associated with the observed rating m on occasion t for teacher i in school j.

(Note, in this data set there is only one observed rating per occasion. As a result the number of

level-1 and level-2 units are identical.)

In most applications, ε_{mtij} is unknown and assumed normally distributed with constant variance. In contrast in this application, the Rasch measurement model for the observed outcomes, Y_{mtij}, also provides a standard error estimate for each observed measure, s_{mtij}. We explicitly represent this by multiplying both sides of the level-1 model by the inverse of the standard error, $a_{mtij} = s_{mtij}^{-1}$, yielding

$$Y^*_{mtij} = a_{mtij}\psi_{0tij} + e^*_{mtij}, \quad e^*_{mtij} \sim N(0,1).$$

The variance at level-1 is now assumed known and fixed at a value of 1.0.

2. Constructing the MDM file from raw data

The user has the same range of options for data input for HLM4 as for HLM3. We will use SPSS file input for the illustrative example.

Data input requires a level-1 file (in our illustration a measurement data file), a level-2 file ("true scores" file), a level-3 (teacher level), and a level-4 (school level) file.

Level-1 file. The level-1 file, MEASURE.SAV, has 1317 observations collected on 219 teachers on up to 9 different occasions. Data for the first three teachers are shown below. Each of these teachers was observed on three occasions. (Some teachers in the study were observed on as many as nine occasions over three years.)

The first column contains the level-4 (*i.e.*, school) ID, next is the level-3 (*i.e.*, teacher) ID, and this is followed by the level-2 (*i.e.*, occasion) ID. We see that the first record comes from school 1100, teacher 1100002, and occasion 11000026. Following the teacher ID fields are that teacher's values on two variables:

- expertis
  A composite Rasch measure of teachers' classroom literacy practice rated on some particular occasion (weighted by the inverse of its standard error of measurement.)
- invstder
  The inverse of the standard error of measurement associated with that individual rating (the standard errors are generated as part of the Rasch rating scale model.)

	schid	tchrid	occasid	expertis	invstderr
1	1100	1100002	11000026	-2.862	4.472
2	1100	1100002	11000027	-1.850	5.000
3	1100	1100002	11000028	-2.182	4.642
4	1100	1100011	11000116	5.750	5.000
5	1100	1100011	11000117	4.105	5.263
6	1100	1100011	11000118	7.150	5.000
7	1100	1100012	11000123	2.227	4.545
8	1100	1100012	11000124	.591	4.545
9	1100	1100012	11000125	2.913	4.348
10	1100	1100013	11000136	.400	5.000

Level-2 file. The level-2 units consisted of the 1317 occasions when measurements on classroom literary practice were made. The data are stored in the file OCCAS.SAV. The level-2 data for the first nine records are listed below. It has the same three ID's as the level-1 file. The two occasion-level variables are included in the file:

- occasion
  This variable identifies the specific data collection time point, counted up from the first study occasion in the fall of year1 (a value of 0) through the end of the study in the spring of year 3 (a value of 8).
- artifact
  A dummy variable introduced into the analysis to adjust for a measurement artifact that occurred with the first-year spring scores (at occasion = 2).

	schid	tchrid	occasid	occasion	artifact
1	1100	1100002	11000026	3.000	.000
2	1100	1100002	11000027	4.000	.000
3	1100	1100002	11000028	5.000	.000
4	1100	1100011	11000116	3.000	.000
5	1100	1100011	11000117	4.000	.000
6	1100	1100011	11000118	5.000	.000
7	1100	1100012	11000123	3.000	.000
8	1100	1100012	11000124	4.000	.000
9	1100	1100012	11000125	5.000	.000
10	1100	1100013	11000136	3.000	.000

The first teacher in this data file, Teacher 1100002 in school 1100, was observed on three occasions during the second year of the study (*i.e.* occasions 3 through 5). The same was true for the next two teachers. In general, the data collection patterns vary among teachers in this study depending upon their employment history at the school and when they first became eligible for classroom coaching.

Level-3 file. The level-3 units are the 219 teachers. The data are stored in the TCHR.SAV file. The

first field is the school ID and the second is the teacher ID. Note that each of the first ten teachers is in school 1100. There are six variables in this file:

- coach
  The average number of one-on-one coaching sessions per month that each teacher received over the course of the study
- newwtch
  A dummy variable indicating that the teacher had three or fewer years of classroom teaching experience at onset of study participation
- pdpart
  A composite measure of teachers' exposure to literacy professional development prior to the onset of the study
- scmt
  A scale score on the teacher's commitment to the school measured at study onset
- y2ent
  A dummy variable indicating the teacher began work at the school during the second year of the study
- y3ent
  A dummy variable indicating the teacher began work at the school during the third year of the study

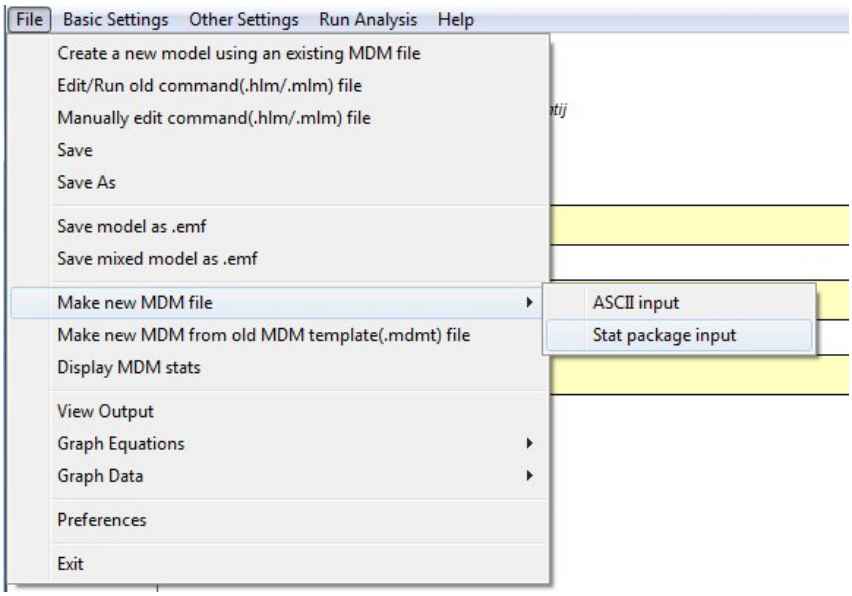
	schlid	tchrid	coach	newtchr	pdpart	scmt	y2ent	y3ent
1	1100.00	1100002	.571	.000	.842	-.292	1.000	.000
2	1100.00	1100011	.571	1.000	-.361	-.813	1.000	.000
3	1100.00	1100012	.755	.000	1.653	.267	.000	.000
4	1100.00	1100013	.571	1.000	1.115	.774	1.000	.000
5	1100.00	1100020	.496	.000	.856	1.150	.000	.000
6	1100.00	1100023	.878	.000	-.248	-1.379	.000	.000
7	1100.00	1100025	.731	.000	-.631	1.150	.000	.000
8	1100.00	1100026	.831	.000	-.248	-.813	.000	.000
9	1100.00	1100027	.736	.000	.307	.164	.000	.000
10	1100.00	1100029	.695	1.000	-.292	-.340	.000	.000

Level-4 file. The school level data from 17 schools appear in SCH.SAV. The first field is the school ID. This is followed by:

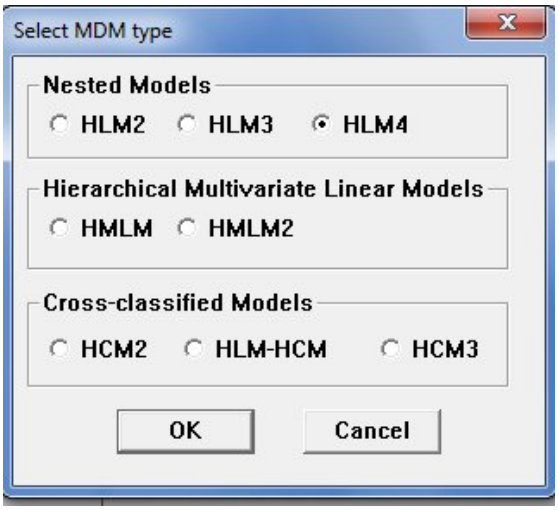
- chgcoach
  A dummy variable indicating that a coaching change occurred during the course of the study. This happened with only one school in the sample.

	schid	chgcoach
1	1100.00	0.000
2	1200.00	0.000
3	1300.00	0.000
4	1400.00	0.000
5	1600.00	0.000
6	1700.00	0.000
7	1800.00	0.000
8	1900.00	0.000
9	2000.00	0.000
10	2100.00	0.000

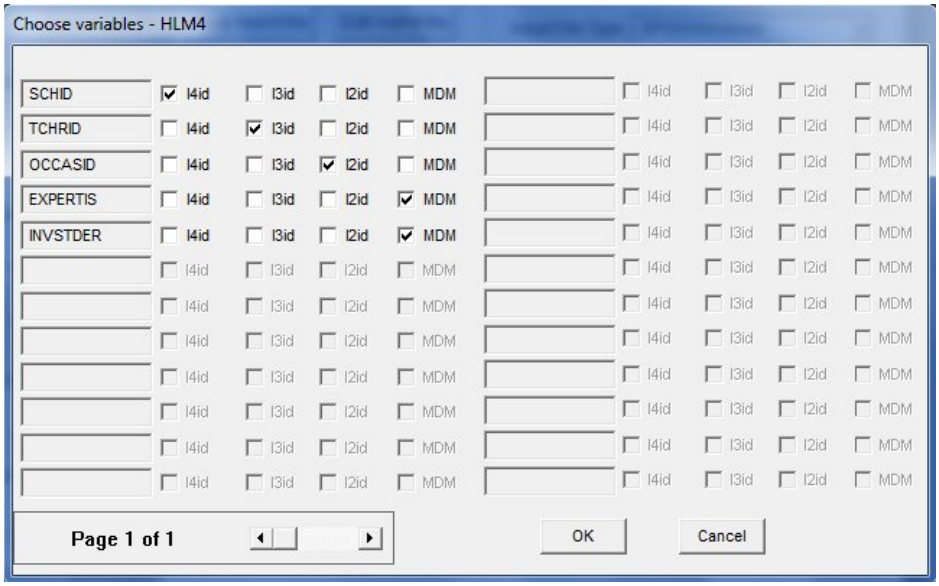
Start by selecting the **File**, **Make new MDM file**, **Stat package** input option from the main menu bar.



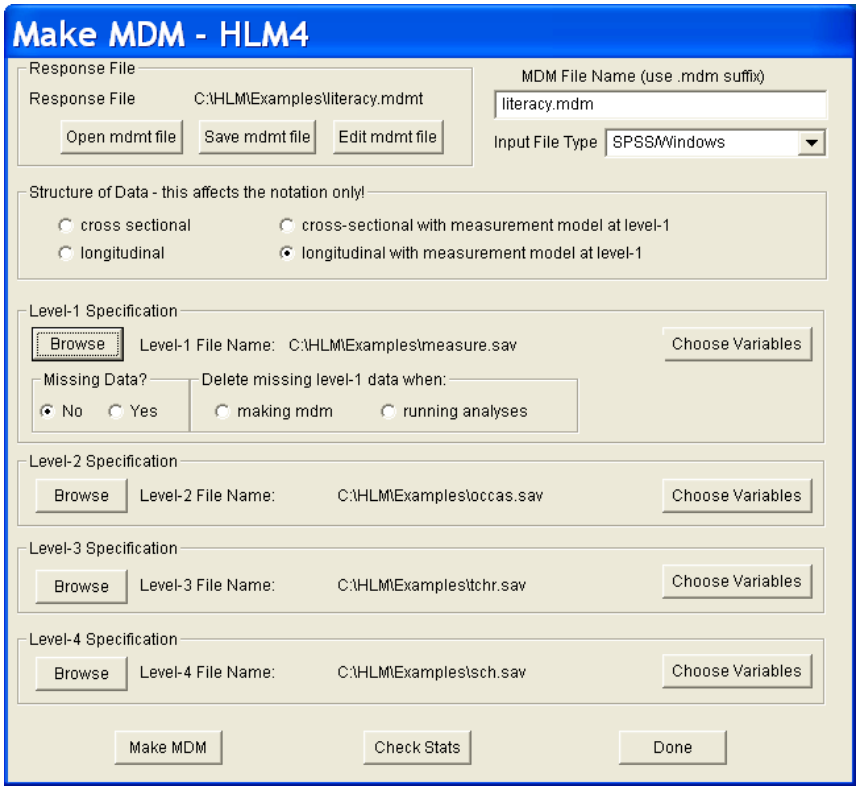
Select the **HLM4** option on the **Select MDM type** dialog box.



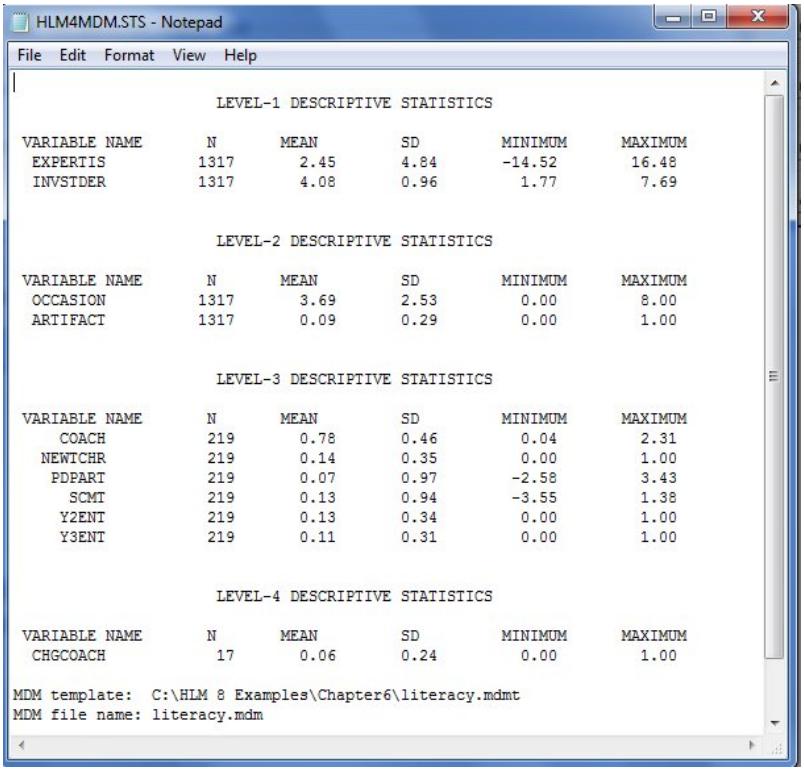
Once the **Make MDM-HLM4** dialog box opens, start by browsing for and selecting variable for inclusion at all levels of the model. Here we are using SPSS, but if another data format is used, the **Files of Type** menu should be used to reflect the type of input file used. The **Choose variables-HLM4** dialog box for selecting level-1 variables is shown below. Note that in the case of a four-level model, three ID variables are required. Proceed to select variables at the other levels in the same way.



The completed Make MDM-HLM4 dialog box is shown below. Input is saved to the file LITERACY.MDMT, which contains a log of the input responses used to create the MDM file, LITERACY.MDM, using MEASURE.SAV, OCCAS.SAV, TCHR.SAV, and SCH.SAV. Saving the file should be the last thing done prior to clicking **Make MDM**. Note that the model notation selected is **longitudinal with measurement model data**. Choosing this option affects the notation used for subscripts and model parameters in the Windows interface and program output.

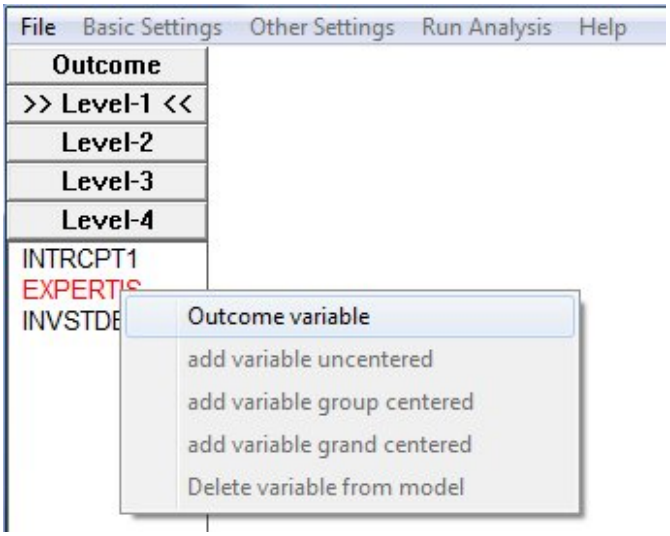


The descriptive statistics for the MDM file, saved to **HLM4MDM.STS**, are shown below.

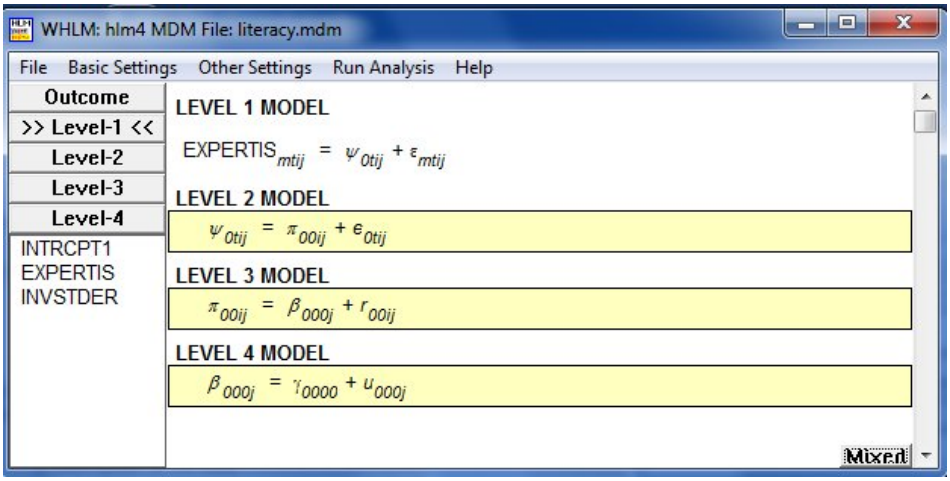


3. Creating the command file

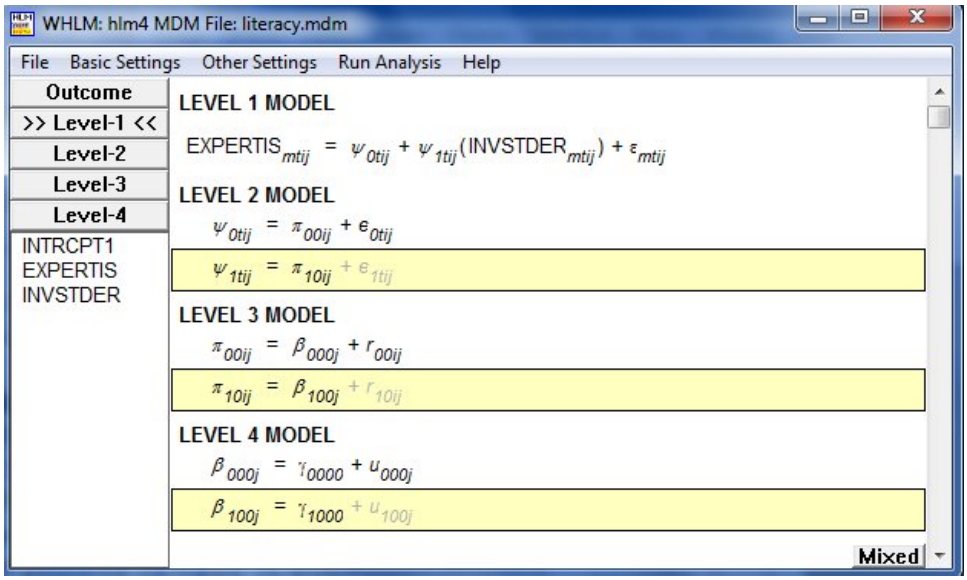
Start by selecting the outcome variable EXPERTIS as shown below.



The fully unconditional model obtained is shown below.

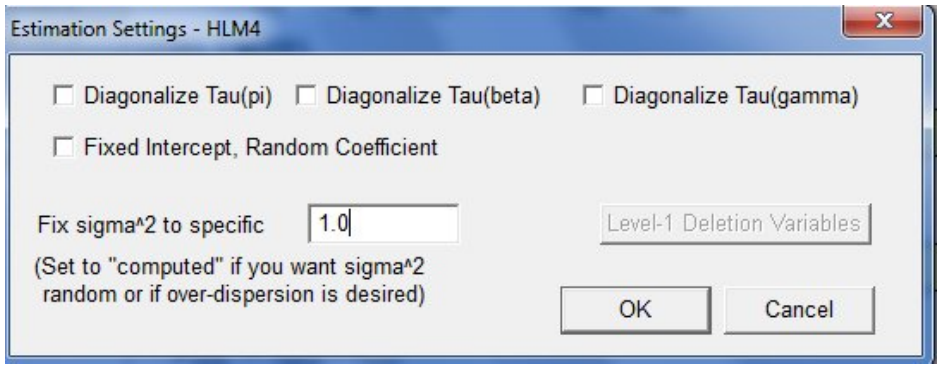


Add the variable INVSTDER by clicking on the name and selectin the **add variable uncentered** option from the pop-up menu to obtain the unconditional model

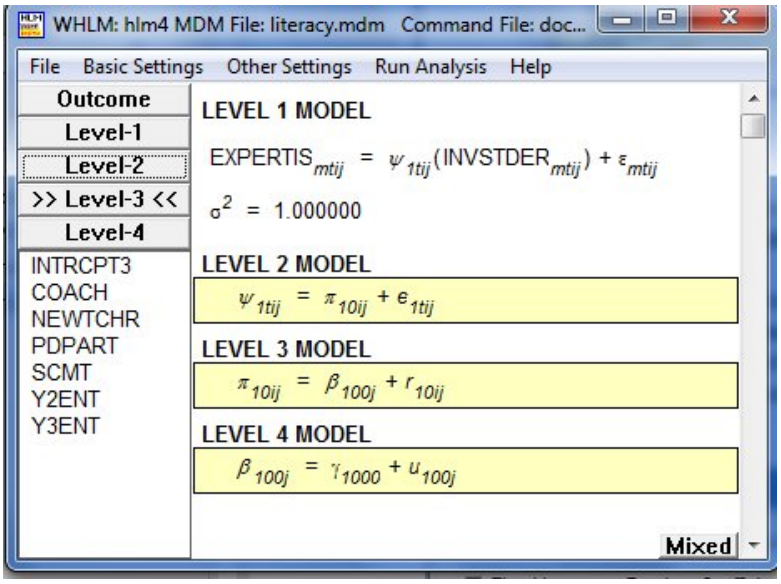


Two additional modifications are required:

1. We need to remove the level-1 intercept. This is done by clicking on **>>Level-1<<** and then on INTRCPT1, selecting the **delete variable from model** option from the pop-up menu.
2. The variance at level-1 is assumed known and fixed at a value of 1.0. To accomplish this, select the **Other Settings**, **Estimation Settings** option to display the **Estimation Settings – HLM4** dialog box. Enter the value "1.0" in the box associated with **Fix sigma^2 to specific**.



Click **OK** to return to the main window and remember to save the command file prior to running the analysis. The final model, complete with fixed sigma, is shown below.



4. Interpreting the output

Output after convergence is shown below.

Final estimation of fixed effects

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INVSTDER, ψ_1 For INTRCPT2, π_{10} For INTRCPT3, β_{100} INTRCPT4, γ_{1000}	0.684973	0.185104	3.700	16	0.002

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

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INVSTDER, e_1	0.95666	0.91520	1080	13451.88848	<0.001

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INVSTDER/INTRCPT2,r_{10}	0.86960	0.75620	201	987.74894	<0.001

Final estimation of level-4 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INVSTDER/ INTRCPT2/INTRCPT3,u_{100}	0.70249	0.49349	16	144.49765	<0.001

Statistics for the current model
Deviance = 7785.342228

Number of estimated parameters = 4

The variation in intercept within teacher ($\text{var}(e_1) = 0.91520$) is higher than between teachers ($\text{var}(r_{10}) = 0.75620$ or between schools ($\text{var}(u_{100}) = 0.70249$). The average intercept is estimated at 0.684973. In the next example, the model will be extended at higher levels to explore the extent to which variables at these levels may impact the relationship between school-based coaching and the development of teachers' expertise in literary instruction.