Biometrical Letters
Vol. 46 (2009), No. 1, 71-86

Hierarchical Linear Models in Education Sciences: an Application

Vítor Valente, Teresa Oliveira

[1]Universidade Aberta and CEAUL, e-mail : vmfvalente@netcabo.pt
[2]Universidade Aberta and CEAUL, e-mail : toliveir@univ-ab.pt

SUMMARY

The importance of hierarchical structured data analysis, based on appropriate statistical models, is very well known in several research areas. In this paper we describe an application in Education Sciences: we have students grouped in classes belonging to schools, which in turn are scattered throughout the country. This grouped organization is labelled as a hierarchical or multilevel structure, and the models usually adopted for statistical analysis of this kind of data are hierarchical linear or multilevel models. The development of these models takes into account data variability within and among the hierarchical levels. We apply a hierarchical linear model (HLM) with two levels – students and schools – in order to identify relevant differences in student performance (10th grade high school in 2004/2005), considering three scientific subjects and comparing two different regions of Portugal: Coastal and Inland.

Key words: hierarchical linear models, multi-level models, multi-level analysis.

1. Introduction

It is well known in the literature that there has been increasing interest in using Hierarchical Linear Models (HLM) to model student (and/or school) academic performance for the purpose of reforming education.

The first researchers in these area employed classical single-level statistical methods, such as linear regression, to model these situations. Nevertheless, when data contain information at more than one level, or when the unit of analysis does not match the unit of randomization in the experiment, then the unit of analysis may become a problem. With classical approaches, one must restrict the data set to eliminate the hierarchy by conducting the analysis at individual or group level. This leads to de-aggregation of the school level data to the individual level or aggregation of the individual level data to the group level, ignoring group identity or individual-level information (Bryk, Raudenbush, 1992).

The limitations of the single-level equation in modelling features, especially for data nested within a group in the form of a hierarchy, led educational researchers to explore an alternative modelling technique, known as hierarchical linear modelling. Such a modelling approach has many advantages to researchers, since there is no need to analyze individual lower (student) level and upper (school) level models separately.

Hierarchical or multilevel data can be analyzed without artificially restructuring the data by employing Multilevel (Goldstein, 1995) or Hierarchical Linear Models (Bryk, Raudenbush, 1992). These models can simultaneously examine effects of both individual and group level variables on an individual level outcome. Moreover, the correlated errors and nonzero ICC (*intra-class correlation* – a basic measure for the degree of dependency in clustered observations) inherent in grouped data are appropriately incorporated in HLM, giving accurate standard error estimates and inferences.

In schools, the more students share common experiences due to closeness in space and/or time, the more similarities they appear to have.

ICC plays an important role in this kind of analysis because it modifies the error variance in traditional linear regression models. This error variance represents the effect of all omitted variables and measurement errors, under the assumption that these errors are unrelated. In traditional linear models the omitted variables are assumed to have a random and not a structural effect – a debatable assumption in the case of data containing clustered observations (Kreft, de Leeuw, 1998).

The multilevel approach is based on relaxing the assumptions depending on the method, algorithms, and software used. In fact, the HLMs are extensions of the linear regression model that relaxes one of the crucial assumptions of the independence of residuals (Snijders, Bosker, 1999).

2. Two-level HLM

We consider two-level hierarchical data structures and follow the notation of Bryk and Raudenbush (1992).

The HLM assumes hierarchical data, with one response variable measured at the lowest level and explanatory variables at all existing levels. Conceptually the model is often viewed as a hierarchical system of regression equations (Hox, 1998). In our work we have data in J groups or contexts (schools), and a different number of individuals (students) n_j in each group. The data do not necessarily have to be balanced (it is not necessary that $n_j = n_k$ for $j \neq k$). At the student level (lowest) we have the dependent variable Y_{ij} and the explanatory variable X_{ij}, and at school level we have the explanatory variable W_j. The double subscript for these variables indicates that the observations are unique for each student i within each school j.

2.1 Model specifications

Thus, in the two-level hierarchical models, we can have separate level-1 regression equations at each of the level-2 units. The level-1 or within-school model can be represented as:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \qquad (1)$$

where Y_{ij} is the outcome for the ith student in the jth school; X_{ij} is the explanatory variable for the ith student in the jth school; β_{0j} is the intercept for the jth school; β_{1j} is the slope for the jth school; and e_{ij} is the random error for the ith student in the jth school from its school's predicted line. The subscripts for the β coefficients in this equation indicate that they can differ for each school j.

Intercepts β_{0j} and slopes β_{1j} are modelled by explanatory variables in the level-2 or between-school models as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \qquad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \qquad (3)$$

where γ_{00} is the estimated intercept when W_j is equal to zero; u_{0j} is the random error for the jth school from the average intercept; γ_{10} is the estimated slope when W_j is equal to zero; and u_{1j} is the random error for the jth school from the average slope. The γ_{01} and γ_{11} are the regression coefficients associated with the effects of the explanatory school level on the student-level structural relationships. Substitution of (2) and (3) in (1) gives:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} W_j + \gamma_{11} W_j X_{ij} + u_{0j} + u_{1j} X_{ij} + e_{ij} \qquad (4)$$

When more than one variable is used at the first or the second level, subscripts such as *p* (*p = 1, 2, …,* P) can be used for the first level and *q* (*q = 1, 2, …,*Q) can be used for the second level. Then (4) becomes the more general equation (Hox, 1998, 2002; Snijders, Bosker, 1999):

$$Y_{ij} = \gamma_{00} + \gamma_{p0} X_{pij} + \gamma_{0q} W_{qj} + \gamma_{pq} W_{qj} X_{pij} + u_{pj} X_{pij} + u_{0j} + e_{ij} \qquad (5)$$

The first part of (5), $\gamma_{00} + \gamma_{p0} X_{pij} + \gamma_{0q} W_{qj} + \gamma_{pq} W_{qj} X_{pij}$, is called the *fixed part* of the model. The second part, $u_{pj} X_{pij} + u_{0j} + e_{ij}$, is called the *random part*. The term $u_{pj} X_{pij}$ can be regarded as a *random interaction between school and X's*.

The specification of error terms at both the student (*e*) and school (*u*) levels allows HLMs to appropriately model the error in grouped data (i.e., nonzero ICC).

The variables *X* and *W* can be modelled in their original, untransformed metric or can be centred (about respective grand means, or *X* about respective group means) (Sullivan *et al.*, 1999).

2.2 Assumptions

The HLM's assumptions are extensions of the linear modelling restrictions required for single level OLS regression (Bryk, Raudenbush, 1992; Snijders, Bosker, 1999).

In our model (equation 1), Y_{ij} is a continuous dependent variable, so we assume that the errors in the level-1 models are normal random variables with mean zero and common variance σ^2 :

$$E(e_{ij}) = 0 \qquad var(e_{ij}) = \sigma^2 \qquad (6)$$

In the level-2 models (equations 2 and 3) we assume that the parameters β_{0j} and β_{1j} are distributed as i.i.d. multivariate normal with means γ_{00} and γ_{10} respectively, and variances τ_{00} and τ_{11} respectively. The covariance of β_{0j} and β_{1j} is denoted τ_{01}. Level-1 and level-2 errors are homogeneous and uncorrelated.

We summarize below the mathematical expressions of the assumptions, which can be found, for example, in Sullivan *et al.* (1999):

$$E(u_{0j}) = 0 \qquad E(u_{1j}) = 0$$
$$E(\beta_{0j}) = \gamma_{00} \qquad E(\beta_{1j}) = \gamma_{10}$$
$$var(\beta_{0j}) = var(u_{0j}) = \tau_{00} \quad \text{and} \quad var(\beta_{1j}) = var(u_{1j}) = \tau_{11} \tag{7}$$
$$cov(\beta_{0j}, \beta_{1j}) = cov(u_{0j}, u_{1j}) = \tau_{01}$$
$$cov(u_{0j}, e_{ij}) = cov(u_{1j}, e_{ij}) = 0$$

2.3 Estimation and Hypothesis Testing in Two-level HLM

Multilevel analysis produces estimates of the fixed effects (γ parameters), the variances and co-variances of the *e* and *u* error terms, known as the variance components.

The estimators generally used in multilevel analysis are Ordinary Least Squares (OLS) and Full or Restricted Maximum Likelihood (FML or RML) estimators (Hox, 2002; Ferrão, 2003). Computing the ML estimates requires an *iterative* procedure. Several algorithms are available to determine these estimates: EM (Expectation–Maximization); Fisher scoring, IGLS (Iterative Generalized Least Squares), and RIGLS (Residual or Restricted IGLS).

Note that when the number of level-2 units is small ($J < 30$) or the data are extremely unbalanced we should be cautious in interpreting the results of significance tests (tests for covariance components and individual random effects in particular). More research needs to be done to determine the robustness of such tests in the presence of small samples and unbalanced data (Sullivan *et al.*, 1999).

If two models are *nested*, which means that a specific model can be derived from a more general model by removing parameters from the general model, we can compare them statistically using their *deviances*. Deviance is defined as $-2 \times \log$ (likelihood). In general, models with a lower deviance fit better than models with a higher deviance.

More detailed discussion of multilevel or HLM procedures can be found in Bryk, Raudenbush (1992), Longford (1993), Goldstein (1995), Kreft, de Leeuw (1998), Snijders, Bosker (1999) and Hox (2002).

3. Application

The data are extracted from a list of questions concerning 10th-grade high-school students in 2004/2005 relating to three scientific subjects and comparing two different regions of Portugal: the Coastal and Inland regions near Lisbon (Figures 1 and 2) and using data from the official site of GIASE – the Council for Inquiry and Evaluation of the Educative System, Ministry of Education.

We apply the hierarchical linear model (HLM) with two levels, students (lower level) and schools (higher level), in order to identify relevant differences in student performance, considering that they are influenced by inherent characteristics of each student and by the environment in which they are placed.

Throughout this work, we use the package MLwiN 2.02, developed and described by Rasbash *et al*. (2004).

Figure 1. Map of Portugal

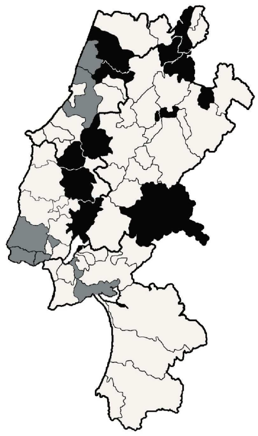

Figure 2. Districts
Grey: Coastal region; Black: Inland region

3.1 Objectives

In order to show that the HLMs are appropriate to identify relevant factors in student performance, we aim to identify whether there are relevant differences in average student performance between schools in the Coastal and Inland regions, which explanatory variables at different levels affect the output variable (average student performance) and how much variability we must have at each output level, and whether the students' distribution by school is random.

3.2 Selected variables

Table 1 shows the selected variables that are used in the construction of the different intermediate models and the final model.

It must be pointed out that the variable MDIDA_T is the aggregation of data of the variable D_IDADE (age) by class and that the variable MHAB_LIT (academic achievement) is the highest level of parents' education.

In terms of school location, we consider an urban school (URBANA), one that is situated in the main district of an urban area, in a city with more than 250 inhabitants per km^2. All other schools, in the urban area of the city not satisfying the aforementioned condition, are in the suburban category (SUB_URB). Other schools are considered as schools in rural areas (RURAL).

3.3 Results and Analysis

In analysis of HLMs a preliminary study with the explanatory variables is made, in order to verify both contribution and significance in future models.

MHAB_LIT, which represents an inherent characteristic of parents, brings down the variability value between schools by about 24.6% (σ^2_{u0} change from 0.043 to 0.057), while the variability value between students only represents 5.7% (σ^2_e change from 0.844 to 0.895). This result, as well as the results for variables D_IDADE (with 24.6% and 6.6%, respectively), REP_ANT (with 24.6% and 7.4%, respectively) and MDIDA_T (with 42.1% and 4.8%, respectively), suggests that students are not distributed by school in a random way.

The variation in the slopes across the schools' summary lines ($\sigma^2_{u1\,\mathbf{s.e.}}$) and the covariance between the school intercepts and slopes ($\sigma^2_{u01\,\mathbf{s.e.}}$) for the explanatory variables D_IDADE – Intercept and MDIDA_T – Intercept are: 0.029 (0.016)*** and -0.024 (0.016) for D_IDADE – Intercept; and, 0.491 (0.210)** and -0.212 (0.098)*** for MDIDA_T – Intercept, respectively. Moreover, the variables A_EMP (with 26.3% and 12.3%); F_TPC (with 38.6% and 8.9%) and UNIVERS (with 36.8% and 14%) also contribute to explaining the differences between schools, as well as between students. This might partly explain the large differences in academic success between students of technological courses and those of scientific-humanistic courses.

Table 1. Description of selected variables

VARIABLE	DESCRIPTION
Level 1 – Student	
ALUNO	Student identification
ZNOTAS_DC	Average student performance – standardised using Z-scores
D_IDADE	Student age difference from expected age for attendance – 15 years (year 1989): 1990; 1989; 1988; 1987; <1987
REGIÃO	What region? Coastal or Inland
CURSO	Type of course: Science/Humanities or Technology
SEXO	Gender: male or female
A_ASSID	Attainment: yes or no
A_PART	Participation: yes or no
A_EMP	Commitment: yes or no
A_DIST	Attention: yes or no
SAN_BAS	Basic sanitation: yes or no
TEL_FIXO	Phone at home: yes or no
COMPUT	Personal computer (at home): yes or no
INTERNET	Internet facilities: yes or no
N_ASSOAL	Number of rooms: <3; 3; 4 or > 4
F_TPC	Homework effectiveness: always; often; occasionally or never
UNIVERS	Following university studies: yes or no
REP_ANT	Second attendance in the same grade: yes or no
EST_ESCO	Regular study at school: yes or no
IMP_ESC	Importance of learning and school for future career: great; some; little or none
AJU_TPC	Homework support: yes, occasionally or no
F_BIBLIO	Use of Resources Center/Library: yes or no
MHAB_LIT	Family scholarship. Scale 0 to 20 years: without basic studies: (0); basic studies: (4) /(6)/(9); secondary studies: (12); university studies: B.S. (17) M.S. or Ph.D.(20)
PARENTAL	Parental family: yes or no
MDIDA_T	Average D_IDADE within the class
Level 2 – School	
ESCOLA	School identification
REGIÃO	What region? Coastal or Inland
URBANA*	School in “urban” area
SUB_URB**	School in “suburban” area
RURAL***	School in “rural” area
LOCALIZ	Type of school location: “urban”; “suburban” or “rural” – reference category

*URBANA – School in an urban area of a city, in the main district of the city, with more than 250 inhabitants per km^2.
**SUB_URB – School in an urban area of a city, in other districts of the city not considered as URBANA.
***RURAL – School in a rural area: neither URBANA or SUB_URB.

Table 2. Individual Models: estimates of the selected variables

VARIABLE	Estimate β_1 (s.e.)	Estimate σ_e^2 (s.e.)	Estimate σ_{u0}^2 (s.e.)	*Deviance*
D_IDADE	- 0.321 (0.032)*	0.836 (0.032)*	0.043 (0.018)**	3718.264*
D_IDADE - Intercept	- 0.345 (0.050)*	0.820 (0.032)*	0.057 (0.024)**	3708.186*
REGIÃO	0.080 (0.112)	0.895 (0.034)*	0.055 (0.022)**	3815.723
CURSO	0.725 (0.058)*	0.804 (0.031)*	0.057 (0.022)**	3669.890*
SEXO	- 0.234 (0.051)*	0.882 (0.034)*	0.056 (0.022)**	3795.663*
URBANA	0.342 (0.093)*	0.895 (0.034)*	0.030 (0.014)**	3805.857**
SUB_URB	- 0.327 (0.118)**	0.894 (0.034)*	0.039 (0.017)**	3809.561**
RURAL	- 0.095 (0.116)	0.895 (0.034)	0.055 (0.022)	3815.556
A_ASSID	0.720 (0.127)*	0.869 (0.033)*	0.051 (0.020)**	3742.025*
A_PART	0.879 (0.084)*	0.829 (0.032)*	0.049 (0.019)**	3658.742*
A_EMP	1.202 (0.087)*	0.785 (0.030)*	0.042 (0.017)**	3593.008*
A_DIST	- 0.719 (0.086)*	0.847 (0.033)*	0.049 (0.020)**	3674.619*
SAN_BAS	0.786 (0.340)**	0.891 (0.034)*	0.056 (0.022)**	3810.878***
TEL_FIXO	0.254 (0.064)*	0.885 (0.034)*	0.053 (0.021)**	3800.418*
COMPUT	0.381 (0.098)*	0.886 (0.034)*	0.053 (0.021)**	3801.216*
INTERNET	0.311 (0.057)*	0.878 (0.034)*	0.049 (0.020)**	3787.217*
N_ASSOAL	0.174 (0.031)*	0.875 (0.034)*	0.054 0.021()**	3771.654*
F_TPC	1.028 (0.086)*	0.815 (0.031)*	0.035 (0.015)**	3680.003*
UNIVERS	0.760 (0.050)*	0.770 (0.030)*	0.036 (0.015)**	3564.270*
REP_ANT	- 0.666 (0.062)*	0.829 (0.032)*	0.043 (0.018)**	3706.471*
EST_ESCO	- 0.256 (0.115)***	0.894 (0.034)*	0.057 (0.022)**	3793.520*
IMP_ESC	0.452 (0.092)*	0.880 (0.034)*	0.053 (0.021)**	3792.398*
AJU_TPC	- 0.201 (0.076)**	0.890 (0.034)*	0.057 (0.022)**	3809.267**
F_BIBLIO	0.203 (0.056)*	0.887 (0.034)*	0.054 (0.021)**	3803.138**
MHAB_LIT	0.054 (0.006)*	0.844 (0.033)*	0.043 (0.018)**	3675.124*
PARENTAL	0.157 (0.062)**	0.891 (0.034)*	0.056 (0.022)**	3809.793**
MDIDA_T	- 0.792 (0.088)*	0.852 (0.033)*	0.033 (0.015)**	3740.012*
MDIDA_T - Intercept	- 0.844 (0.177)*	0.819 (0.032)*	0.103 (0.050)**	3702.244*

* Significant to $\alpha \leq 0.001$ ** Significant to $\alpha \leq 0.01$ *** Significant to $\alpha \leq 0.05$

All values not marked with an asterisk are not significant to $\alpha \leq 0.05$.

It is worth noting that the variables concerning the location of schools – URBANA and SUB_URB – also explain the considerable differences among schools. Students from URBANA area schools have a better achievement rate than those from SUB_URB.

We choose to aggregate some explanatory variables in intermediate models to test their significance as a group. The results are presented in Table 3.

Table 3. Results of Two-level Hierarchical Linear Model

	Unconditional Model Random Intercept	Model I	Base Model	Base Model Random Slope	Model II	Model III
Parameters	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)
FIXED						
Intercept	-0.011 (0.057)	-0.301 (0.092)*	-0.074 (0.100)	-0.059 (0.099)	-1.015 (0.161)*	-1.148 (0.356)*
D_IDADE		-0.244 (0.031)*	-0.249 (0.031)*		-0.193 (0.030)*	-0.220 (0.031)*
D_IDADE – Intercept				-0.256 (0.039)*		
CURSO		0.584 (0.059)*	0.259 (0.093)*	0.249 (0.093)**	0.211 (0.088)**	0.240 (0.092)**
SEXO		-0.117 (0.049)*	-0.107 (0.049)**	-0.114 (0.049)**	-0.069 (0.049)	-0.133 (0.048)**
URBANA		0.165 (0.104)	-0.309 (0.138)**	-0.311 (0.133)**	-0.377 (0.133)**	-0.328 (0.135)**
SUB_URB		-0.161 (0.120)	-0.429 (0.144)*	-0.431 (0.136)*	-0.349 (0.140)**	-0.420 (0.140)**
CURSO × URBANA			0.632 (0.135)*	0.622 (0.135)*	0.623 (0.126)*	0.616 (0.133)*
CURSO × SUB_URB			0.378 (0.138)*	0.368 (0.139)**	0.285 (0.129)***	0.356 (0.136)**
A_ASSID					0.298 (0.113)**	
A_PART					0.617 (0.077)*	
A_EMP					0.677 (0.090)*	
A_DIST					-0.337 (0.080)*	
SAN_BAS						0.524 (0.331)
TEL_FIXO						0.125 (0.063)***
COMPUT						0.159 (0.097)
INTERNET						0.121 (0.061)***
N_ASSOAL						0.083 (0.030)**
RANDOM						
Level 2: Schools – Intercept	0.057 (0.022)**	0.032 (0.014)***	0.023 (0.011)***	0.033 (0.016)***	0.025 (0.011)***	0.019 (0.01)***
Level 2 Schools – Slope				0.011 (0.009)		
Level 2: Schools – Interaction				-0.014 (0.010)		
Level 1: Students	0.895 (0.034)*	0.768 (0.029)*	0.758 (0.029)*	0.752 (0.029)*	0.639 (0.025)*	0.740 (0.028)*
-2 log(likelihood)	3816.218	3596.639	3574.844	3570.928	3269.826	3525.414
Number of valid data	1387	1387	1387	1387	1357	1382

* Significant to α ≤ 0.001   ** Significant to α ≤ 0.01   *** Significant to α ≤ 0.05

All values not marked with an asterisk are not significant to α ≤ 0.05.

Table 3 (cont.). Results of Two-level Hierarchical Linear Model

	Model IV	Model V	Final Model	Final Model Random Slope	Final Model Random Slope with Interactions
Parameters	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)
FIXED					
Intercept	-0.502 (0.133)*	-0.353 (0.139)**	-1.369 (0.175)*	-1.363 (0.175)*	-1.777 (0.190)*
D_IDADE	-0.046 (0.039)	-0.209 (0.032)*	-0.055 (0.040)		
D_IDADE – Intercept				-0.049 (0.048)	-0.037 (0.047)
CURSO	-0.096 (0.101)	-0.020 (0.138)	0.140 (0.098)	0.136 (0.098)	-0.164 (0.139)
SEXO	-0.193 (0.093)***	-0.165 (0.049)*	-0.072 (0.046)	-0.077 (0.046)	-0.196 (0.091)***
URBANA	-0.428 (0.116)*	-0.346 (0.131)**	-0.403 (0.125)*	-0.391 (0.123)	-0.366 (0.119)**
SUB_URB	-0.415 (0.118)*	-0.239 (0.174)	-0.383 (0.127)**	-0.374 (0.125)	-0.177 (0.158)
CURSO × URBANA	0.641 (0.125)*	0.520 (0.134)*	0.523 (0.122)*	0.511 (0.123)	0.453 (0.124)*
CURSO × SUB_URB	0.338 (0.130)**	0.292 (0.142)***	0.224 (0.125)***	0.222 (0.127)	0.172 (0.132)
CURSO × SEXO	0.231 (0.107)***				0.153 (0.104)
A_ASSID					
A_PART			0.417 (0.077)*	0.418 (0.077)	0.416 (0.077)*
A_EMP			0.551 (0.091)*	0.537 (0.091)	0.522 (0.091)*
A_DIST			-0.291 (0.079)*	-0.295 (0.079)	-0.289 (0.079)*
SAN_BAS					
TEL_FIXO			0.134 (0.055)**	0.138 (0.055)	0.144 (0.055)**
COMPUT					
INTERNET					
N_ASSOAL					
F_TPC	0.430 (0.122)*		0.291 (0.087)*	0.299 (0.087)	0.307 (0.086)*
UNIVERS	0.168 (0.108)		0.303 (0.052)*	0.305 (0.052)	0.311 (0.052)*
REP_ANT	- 0.292 (0.076)*		-0.233 (0.075)*	-0.241 0.075	-0.253 (0.075)*
EST_ESCO	-0.171 (0.099)***				
IMP_ESC	0.214 (0.083)**				
AJU_TPC	-0.269 (0.066)*		-0.318 (0.064)*	-0.322 (0.064)	-0.313 (0.064)*
F_BIBLIO	0.088 (0.050)***		0.112 (0.048)**	0.116 (0.048)	0.110 (0.048)***
F_TPC × UNIVERS	0.460 (0.157)**				
MHAB_LIT		0.022 (0.012)***	0.031 (0.005)*	0.031 (0.005)	0.014 (0.011)
PARENTAL		0.136 (0.057)**	0.084 (0.053)	0.081 (0.053)	0.083 (0.053)
CURSO × MHAB_LIT		0.028 (0.013)***			0.027 (0.012)
MHAB_LIT × SUB_URB		-0.022 (0.012)***			-0.016 (0.011)
MDIDA_T			0.181 (0.104)	0.179 (0.105)	0.178 (0.104)
RANDOM					
Level 2: Schools – Intercept	0.006 (0.005)	0.016 (0.009)***	0.016 (0.008)***	0.020 (0.010)***	0.016 (0.009)***
Level 2 Schools – Slope				0.012 (0.009)	0.011 (0.009)
Level 2: Schools – Interaction				-0.008 (0.008)	-0.008 (0.007)
Level 1: Students	0.653 (0.025)*	0.729 (0.028)*	0.587 (0.023)*	0.581 0.023*	0.578 (0.023)*
-2 log(likelihood)	3300.064	3463.731	3072.828	3069.112	3058.551
Number of valid data	1364	1366	1324	1324	1324

* Significant to $\alpha \leq 0.001$ ** Significant to $\alpha \leq 0.01$ *** Significant to $\alpha \leq 0.05$
All values not marked with an asterisk are not significant to $\alpha \leq 0.05$.

The observed *intra-class correlation* (ICC) is:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} = \frac{0.057}{0.057 + 0.895} = 0.0598 \cong 6\% . \tag{8}$$

This result is a rather low one, compared with other results of educational researchers – values between 0.05 and 0.20 are common (Snijders, Bosker, 1999). This indicates that the grouping according to schools leads to a low similarity between the results of different students in the same school, although within-school differences are far larger than between-school differences.

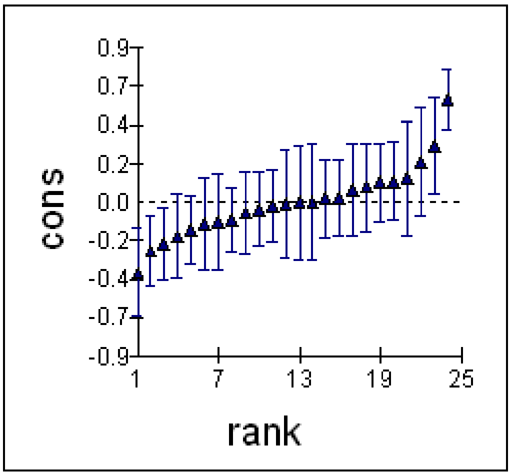

Figure 3. 24 level-2 residuals (output of MlwiN 2.02)

In this caterpillar plot we have 24 level-2 residuals plotted, one for each school in the data set. Looking at the confidence intervals around them, we can see a group of 5 schools (3 at the lower end and 2 at the upper end of the plot) where the confidence intervals for their residuals do not overlap with zero.

These residuals represent school random errors from the overall average predicted by the fixed parameter, γ_{00}; this means that these are a few schools that differ significantly from the average, at the 5% level.

Estimated residuals, at any level, can be used to check model assumptions. One such assumption is that the residuals at each level follow Normal distributions. This assumption may be checked using a Normal probability plot, in which the ranked residuals are plotted against corresponding points on a Normal distribution curve. If the Normality assumption is valid, the points on a Normal plot should lie approximately on a straight line (Rasbash *et al*. 2004).

The plots (figures 4 and 7) looks fairly linear, which suggests that the assumption of Normality is reasonable. This is not surprising in this case since our response is nearly Normal.

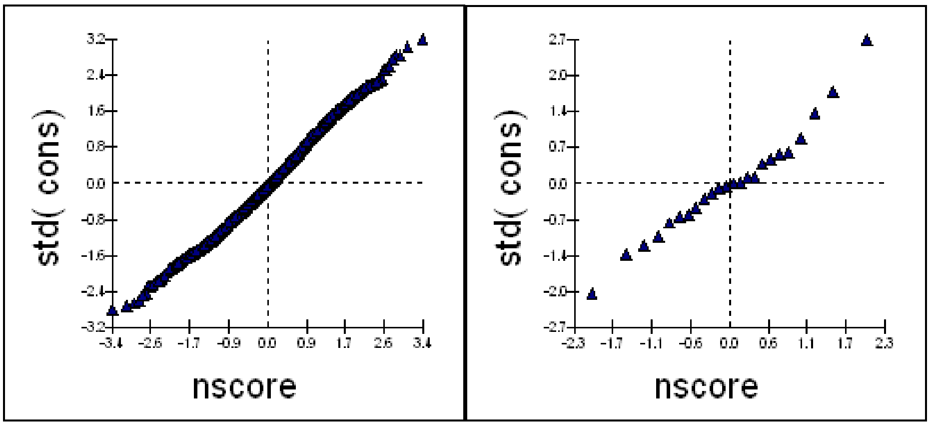

Figure 4. Plot of student and school residuals for unconditional model (output of MlwiN 2.02)

It seems there are no evident differences between Coastal and Inland schools, as we can see in the model of Figure 4. REGIAO's coefficient is less than its standard error, and therefore not statistically significant.

$$\text{ZNOTAS\_DC}_{ij} = \beta_{0j} + 0.080(0.112)\text{REGIAO}_j + e_{ij}$$

$$\beta_{0j} = -0.052(0.081) + u_{0j}$$

$$u_{0j} \sim \text{N}(0,\ \sigma^2_{u0}) \quad \sigma^2_{u0} = 0.055(0.022)$$

$$e_{ij} \sim \text{N}(0,\ \sigma^2_e) \quad \sigma^2_e = 0.895(0.034)$$

$$-2*loglikelihood = 3815.723(1387 \text{ of } 1387 \text{ cases in use})$$

Figure 5. HLM for the REGIAO variable (output of MlwiN 2.02)

The base model is formed by variables representing inherent characteristics such as D_IDADE (age of the student), CURSO (the course chosen by the student), SEXO (student's gender), URBANA and SUB_URB (location of the school). This model explains 59.6% of the existing variability among schools and 15.3% of variability among students.

Now, considering the model with random slopes, we can conclude that, although the coefficient is significant, an improvement of the model is not evident. Yet it can be stated that schools do not bring about a large difference between students of different ages, that is, they are considered to be "equitable" according to Bryk, Raudenbush (1992).

Model II deals with students' attitudes. This is a very significant model, since it contributes with 56.1% and 28.6% to the explanation of the difference between schools and between students, respectively.

Model III presents variables related to some of the possessions and services available at a student's home. This, to a certain extent, is related to the socio-economic standing of the family. The model is very significant, but only the variable TEL_FIXO appears in the final model. Its contribution is 66.7% and 17.3% to the explanation of differences between schools and between students, respectively.

Model IV is composed of variables that show the attitudes and expectations of the students towards the schools and their studies. It is quite significant and the variability between schools is practically explained by these variables – the coefficient is no longer significant. The variability between students is explained to a level of 27%.

In Model V there are variables related to family characteristics: MHAB_LIT (the best academic achievement of the parents) and PARENTAL (type of family). There are also some interactions. The variability between schools is explained to a level of 71.9%, and that between students to 18.5%.

The final model (only random intercepts) is presented in Figure 6.

$$\begin{aligned}\text{ZNOTAS\_DC}_{ij} = {} & \beta_{0j} + \beta_1\text{D\_IDADE}_{ij} + \beta_2\text{CURSO}_{ij} + \beta_3\text{SEXO}_{ij} + \beta_4\text{URBANA}_j + \beta_5\text{SUB\_URB}_j + \\ & \beta_6\text{CURSO.URBANA}_{ij} + \beta_7\text{CURSO.SUB\_URB}_{ij} + \beta_8\text{MHAB\_LIT}_{ij} + \beta_9\text{F\_TPC}_{ij} + \\ & \beta_{10}\text{UNIVERS}_{ij} + \beta_{11}\text{A\_PART}_{ij} + \beta_{12}\text{A\_EMP}_{ij} + \beta_{13}\text{A\_DIST}_{ij} + \beta_{14}\text{REP\_ANT}_{ij} + \\ & \beta_{15}\text{AJU\_TPC}_{ij} + \beta_{16}\text{F\_BIBLIO}_{ij} + \beta_{17}\text{TEL\_FIXO}_{ij} + \beta_{18}\text{MDIDA\_T}_{ij} + \beta_{19}\text{PARENTAL}_{ij} + e_{ij}\end{aligned}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim \text{N}(0, \sigma^2_{u0})$$
$$e_{ij} \sim \text{N}(0, \sigma^2_e)$$

*-2*loglikelihood* = 3072.828(1324 of 1387 cases in use)

Figure 6. Final Hierarchical Linear Model (output of MlwiN 2.02)

Figure 7 shows the student and school residuals.

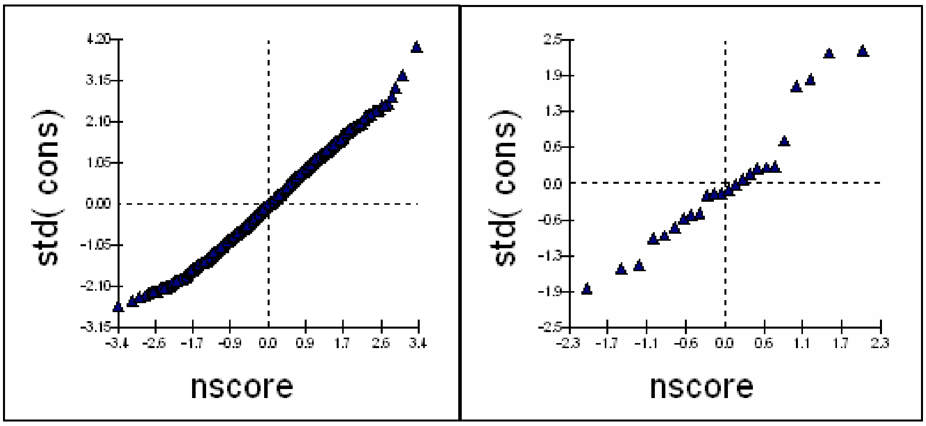

Figure 7. Plot of student and school residuals for final model (output of MlwiN 2.02)

4. Conclusions

Our research provided evidence or the following statements:

1. students with age-grade imbalance have a tendency to score poorly;
2. male students perform worse than female students;
3. students of sciences and humanities perform better than students of technological studies;
4. relative to students from schools in "rural" regions – the reference:
   - students from "urban" schools have better performance than the "rural" ones;
   - students from "suburban" schools have poorer performance than the "rural" ones.

The final model decreases the variability between schools by about 71.9% and that between students about 34.4%. Comparing the final model with the unconditional model we can observe that the value of –2log(likelihood) has decreased from 3816.218 to 3072.828, a difference of 743.39. The change in the deviance value has a chi-squared distribution, 19 degrees of freedom, under the null hypothesis. We therefore conclude that the change is very highly significant, confirming the better fit of the more elaborate model to the data.

References

Bryk A., Raudenbush S. (1992): Hierarchical Linear Models: applications and data analysis methods. London: Sage Publications.

Ferrão M.E. (2003): Introdução aos modelos de regressão multinível em educação. Campinas: Komedi.

Goldstein H. (1995): Multilevel Statistical Models. 2nd Edition. London: Edward Arnold; New York: Wiley.

Hox J. (1998): Multilevel modeling: when and why? In: I. Balderjahn, R. Mathar, M. Schader (Eds.). Classification, data analysis, and data highways (147-154). New York: Springer Verlag. http://www.fss.uu.nl/ms/jh/publist/whenwhy.pdf

Hox J. (2002): Multilevel Analysis: Techniques and Applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Kreft I., de Leeuw J. (1998): Introducing Multilevel Modeling. London: Sage Publications.

Longford N.T. (1993): Random coefficient models. New York: Oxford University Press.

MLwiN, Version2.02 (2005): Multilevel Models Project. Institute of Education. Software package authors: Rasbash, J.; Browne, W.; Healy, M.; Cameron, B. and Charlton, C.

Rasbash J., Steele,, F., Browne W., Prosser B. (2004): A User's Guide to MLwiN: version 2.0. London: Centre for Multilevel Modelling, Institute of Education, University of London.

Snijders T., Bosker R. (1999): Multilevel Analysis: An introduction to basic and advanced multilevel modeling. London: Sage Publications.

Sullivan L.M., Dukes K.A., Losina E. (1999): Tutorial in Biostatistics: an introduction to hierarchical linear modelling. Statistics in Medicine 18: 855-888.

http://stat.gamma.rug.nl/snijders/sullivan_tutorial.pdf.