MODELLING FACTORS THAT INFLUENCES ACADEMIC ACHIEVEMENT USING MULTILEVEL MODELLING
Siti Aishah Sh.Abdullah & Siti Fatimah Othman
Faculty of Science Computer & Mathematics
University Technology MARA
sitiaishah@kelantan.uitm.edu.my

ABSTRACT

This paper discusses two level multilevel analysis to examine the factors that influences secondary students academic achievement. A total of 7000 secondary students clustered in 31 secondary schools national were selected at random in one of the largest state in Malaysia. Dependent variable used was aggregated merit score based on subjects taken in Malaysian lower secondary national examination score. The first level is the within students level and second level is between schools. School effects on student achievement were estimated using factors such as school location, percent of students that received financial aid, percent male students and type of schools. The finding suggested that there were significant differences in students performances between schools. School intraclass correlation explained about 38.7% of variations in students performances are due to differences in schools. Gender showed an inverse relationship with academic grade score while school location indicates small influential effect. There is significant interaction effect between gender and type of schools in explaining academic achievement in this study. Implications are discussed.

Field of Research: Hierarchical linear modelling, regression, academic achievement

Introduction

School effects studies usually focused on analyzing the associations between school contextual factors and students achievement. School characteristics such as school region, school types, school enrolment and poverty percentage are important correlates of school effectiveness (Goldstein, 1995). In many of school effect research effort was made to model achievement as a function of school context and students background estimated at school level or adjusted for students level. Typical multiple regression have been extensively used in many schools effect researches. When using multiple regression methods schools and students level predictors are introduced simultaneously at the students level or aggregated to group levels. The analysis is then conducted at the individual level or group levels separately. In such cases the analysis do not take into account the clustering nature of the data and hence typical regression model cannot estimates data in school variation. Hierarchical data such as students achievement in several group of schools ought to be treated appropriately since ordinary regression analysis only indicated a relatively small impact of school characteristics on students (Goldstein,1995;Hox,1994).
Educational data often are organized at students, school or district levels. Students from similar school background tend to exhibit similar characteristics compared to students from another background. Hierarchical data as group data is seen to be more appropriate.

1. Analysis of school effect

A multilevel analysis that explicitly models students structure in which students are grouped within schools has many advantages in providing statistically more accurate and efficient estimates of student achievement model. There are various models that have been proposed for multilevel or clustered data such as this. Hierarchical Linear Models (HLM) is an estimation method that was developed by (Hox,1999) is growing in popularity in research areas such as health and psychology, economy and social studies[2],[3].Ordinary multiple regression have been extensively used in many schools effect researches. When using multiple regression methods schools and students level predictors are introduced simultaneously at the students level or aggregated to group levels. The analysis are usually then conducted at the individual level or group levels separately. In such cases these analysis do not take into account the clustering nature of the data and hence typical regression model cannot give accurate estimations of data in school variation. In contrast HLM software take clustering of data of students within schools and permit computation of school variance simultaneously (Sneider & Beckers,1999). In this paper HLM is used to treat students within a single schools as unit of observations at first level and their schools as the second level of analysis Both individuals within schools are treated as random samples from their respective populations at the fist level.

2. Research objectives

The first objective of this study is apply multilevel analysis method to model school performances of Malaysian students by taking into consideration school characteristics in addition to students background simultaneously. Second objective is to estimate school effects in explaining students’ achievement.

3.1 Research questions

Specifically there are three main questions addressed in this study.
   i. What is significant factors that explain academic achievement of Malaysian students?
   ii. How much variances is explained by school in students academic achievement?
   iii. Which school variables significantly contributed towards students academic achievement?

4 Methodology

4.1 Data and variables

Data related to all students achievement in secondary examination grade in year 2010 were obtained from state education department. These include all subjects score and total average score or merit score for each students (NGP) in every schools.
Merit score systems are calculated by Malaysian education department based on aggregated score of all subjects taken by students. Score ranges from excellent grade (18) until zero. The higher the number the better the merit of students achievement. Table 1 showed all variables used at students and school levels.

4.2 Method of Analysis

Fifty schools were randomly selected and 100 students were again randomly selected from each of these schools making a total of 7985 students. Missing students information were excluded from the study. The data was separated between students and schools. All data was initially input into SPSS17.0 and later HLM 6 was used to create two level models.

Table 1. Factors at each Hierarchical level that explains academic achievement.

<table>
<thead>
<tr>
<th>Data level</th>
<th>Example</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2</td>
<td>School</td>
<td>School location LOC(urban=0,rural=1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>School TYPE(national type=0, non national type =1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>School % poverty rate (POV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>School % male students (PMALE)</td>
</tr>
<tr>
<td>Level 1</td>
<td>Students</td>
<td>Gender (Male=1,female=0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Race (Malays= 1, non malay = 0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NGP* (grade 0-18)</td>
</tr>
</tbody>
</table>

* aggregated merit score - the outcome variable

4.2.1 HLM Model

This study used two level HLM to explore school variability and the effect on average students achievement (NGP) based on merit score system formula used by Malaysian education department. The simplest model initially is null model. This model do not have predictors at either level of the analysis. Level 1 of the model is

\[ Y_{ij} = \beta_{ij} + r_{ij} \]  

Let i indicates the individual and the j schools levels. \( \beta_{ij} \) is the mean for school j and \( r_{ij} \) is the deviation of each individual observations (score) from the average school mean. School intercept are treated as randomly varying. The second level model is:

\[ \beta_{0j} = \gamma_{00} + \mu_{0j} \]  

where the mean for the school, \( \beta_{0j} \) is a function of grand mean \( \gamma_{00} \) and \( \mu_{0j} \) is a random error of the difference between school mean and grand mean of outcome variable.

Combining both of these equations simultaneously we now obtained the model:

\[ Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij} \]  

This analysis will give the estimates of the grand mean and the estimates of variances within individuals and between groups. Additionally important outcome of the first null model is estimating the intraclass coefficient. This coefficient is the proportion of
outcome variance due to differences between schools characteristics. The next conditional models HLM model is developed by taking into account more predictors from individual and also from school levels. These models include selected variables as predictors to test each main effect and interaction effect by considering fixed effect and random effect in each of the models.

5. Results

5.1 Null Model

The null model is basically the one way ANOVA model. This model provides the variations in academic score within and between schools and realibility of each school sample mean as estimates of true population mean. Table 2 indicates the analysis for null model.

\[
\begin{align*}
\text{Level 1: } \text{NGP}_{ij} &= \beta_{0j} + r_{ij} \\
\text{Level 2: } \beta_{0j} &= \gamma_{00} + u_{0j} \\
\text{Mixed Model: } \text{NGP}_{ij} &= \gamma_{00} + u_{0j} + r_{ij} \tag{4}
\end{align*}
\]

Table 2 fixed effect result, indicates that the mean estimates for academic achievement is 2.63 with standard error of 0.162 and 95% confidence interval of \([2.63 \pm 1.96(0.162)]\) which is a plausible range for the means. Also we notice that the variance indicates that there is a significant differences between the mean \([\chi^2=6992.996, p<0.001]\). In addition an estimator of sample mean value of 0.993 gives reliable estimates of mean population value.

Table 2. Unconditional Model estimates of fixed and random components

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>Approx. d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>For INTRCPT1, (\beta_0)</td>
<td>8.006631</td>
<td>0.439712</td>
<td>18.209</td>
<td>34</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>INTRCPT2, (\gamma_{00})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Effect</td>
<td>Standard Deviation</td>
<td>Variance Component</td>
<td>d.f.</td>
<td>(\chi^2)</td>
<td>p-value</td>
</tr>
<tr>
<td>INTRCPT1, (u_0)</td>
<td>2.62682</td>
<td>6.90017</td>
<td>34</td>
<td>3516.61496</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>level-1, (r)</td>
<td>3.31180</td>
<td>10.96803</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the values in Table 2, the intraclass correlation, i.e the proportion of between schools variance can be calculated,

\[
\rho = \frac{\tau_{00}}{[\tau_{00} + \sigma^2]} = \frac{6.90017}{[6.90017 + 10.96803]} = 0.38617.
\]

This indicates that nearly 38.6% of variability in academic achievement can be contributed to schools differences or contributions from second level. Therefore we established that hierarchical model is necessary.
5.2 Model with student predictor

This model contains predictors gender which is from students level only. There is no predictor at school level.

Level 1: \( NGP_{ij} = \beta_0j + \beta_{1j} \times (\text{gender}_{ij}) + r_{ij} \)

Level 2: 
\[
\begin{align*}
\beta_0j &= \gamma_{00} + u_{0j} \\
\beta_{1j} &= \gamma_{10}
\end{align*}
\]

Mixed Model 1 \( NGP_{ij} = \gamma_{00} + \gamma_{10} \times \text{gender}_{ij} + u_{0j} + r_{ij} \)

(5)

The reliability of this estimate is 0.995. This result shows that gender is significant factors in explaining differences in students academic achievement. Since this value is negative NGP value exhibit inverse relation with male students.

Next analysis of intraclass effect for model with gender at students level is calculated as follows,

\[ \rho_1 = \frac{\sigma^2_{null} - \sigma^2_{random}}{\sigma^2_{null}} = \frac{[10.16930 - 6.7013]}{10.16930} = 0.341026 \]

The intraclass value with gender effect is lower indicating that the effect of gender between different schools are too small. At students level differences in academic achievement due to gender differs significantly (\( \chi^2=3642.82314, p<0.001 \)). They explained almost 34% of differences in NGP score. Students race is not significant and therefore do not contribute to any changes in students level variations in this model. The estimated fixed effect (\( \gamma_{10} = -1.884164 \)) at student level however indicate a strong impact favouring female student.

5.3 Full Model

The next model is a full model containing all students and school predictor. Only students socioeconomic status (SES) level and school poverty (POV) at school level were found significant.

**Table 3.** Estimates of fixed and Random effect in unconditional model

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-ratio</th>
<th>Approx. d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>For INTRCPT1, ( \beta_0 )</td>
<td>INTRCPT2, ( \gamma_{00} ) = 8.011438</td>
<td>0.441152</td>
<td>18.160</td>
<td>34</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>For GENDER slope, ( \beta_i )</td>
<td>INTRCPT2, ( \gamma_{10} ) = -1.884164</td>
<td>0.082489</td>
<td>-22.841</td>
<td>6619</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Variance Component</th>
<th>d.f.</th>
<th>( \chi^2 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRCPT1, ( u_0 )</td>
<td>2.59810</td>
<td>6.75013</td>
<td>34</td>
<td>3642.82314</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The model with SES and POV are shown.

Level 1: \( NGP_{ij} = \beta_0j + \beta_{1j} \times (\text{gender}_{ij}) + r_{ij} \)
Level 2: \[ \begin{align*} 
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{schooltype}) + \gamma_{02}(\text{location}) + \gamma_{03}(\text{malepercent}) + u_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{schooltype}) + \gamma_{12}(\text{location}) + \gamma_{13}(\text{malepercent}) + u_{1j} 
\end{align*} \]

Mixed Model: \[ \begin{align*} 
\text{NGP}_{ij} &= \gamma_{00} + \gamma_{01}(\text{schooltype}) + \gamma_{02}(\text{location}) + \gamma_{03}(\text{malepercent}) + \gamma_{10}(\text{gender}) + \gamma_{11}(\text{schooltype})(\text{gender}) + \gamma_{12}(\text{location})(\text{gender}) + \\
&\quad \gamma_{13}(\text{malepercent})(\text{gender}) + u_{0j} + u_{1j}(\text{gender}) + r_{ij} 
\end{align*} \] (6)

The analysis shown in Table 4, indicated that achievement of students differ significantly according to type of schools and location. School type favours non national type of school while urban schools indicates higher scores compared to rural.

| Table 4 The estimates of fixed an random effect of full model |
|-------------------|-------------------|-----------------|--------|--------|
| Fixed Effect      | Coefficient       | Standard error  | t-ratio| Approx. |
|                   |                   |                 | d.f.   | p-value|
| For INTRCPT1, \(\beta_0\) |                   |                 |        |        |
| INTRCPT2, \(\gamma_00\) | 8.008687          | 0.227751        | 35.164 | 31     | <0.001 |
| \(\text{schooltype}\), \(\gamma_01\) | 2.924251          | 0.360353        | 8.115  | 31     | <0.001 |
| \(\text{location} \), \(\gamma_02\) | 1.358129          | 0.430027        | 3.158  | 31     | 0.004  |
| \(\text{malepercent} \), \(\gamma_03\) | 0.004120          | 0.014617        | 0.282  | 31     | 0.780  |
| For GENDER slope, \(\beta_1\) |                   |                 |        |        |
| INTRCPT2, \(\gamma_{10}\) | -1.636875         | 0.114998        | -14.234| 31     | <0.001 |
| \(\text{schooltype} \), \(\gamma_{11}\) | 1.328290          | 0.291784        | 4.552  | 31     | <0.001 |
| \(\text{location} \), \(\gamma_{12}\) | 0.119940          | 0.202050        | 0.594  | 31     | 0.557  |
| \(\text{malepercent} \), \(\gamma_{13}\) | -0.060967         | 0.028316        | -2.153 | 31     | 0.039  |
| Random Effect     | Standard Deviation| Variance Component| d.f. | \(\chi^2\) | p-value|
| INTRCPT1, \(u_0\) | 1.32471           | 1.75486         | 27     | 864.22515 | <0.001 |
| GENDER slope, \(u_1\) | 0.33704           | 0.11359         | 27     | 40.72438 | 0.044  |
| level-1, \(r\)    | 3.17738           | 10.09573        |        |         |        |

Furthurmore there is significant interaction effect between gender and school type and male percentage in schools. Non national type schools shows a higher merit score of 2.9 points compared to national type school. While interaction effect between gender and schooltype indicated gender effect is significantly higher in national type compared to non national type.
6. Discussion

HLM models shows that there are schools and individual variability in explaining students academic achievement. In this study school variations explains about 38.6 % of students academic score. Students gender were found to be strongly significant and related with academic score. While impact school factors that have significant effect on students merit score is school location and male percentage in the school. Only one variable at individual level is used in this study i.e students gender. Gender account for 34% of differences in academic merit score in this sample of Malaysian population. Similar result was also obtained by Veenstra & Kuyper (2004). Furthmore students social economic status were found to be significant and related with academic score. While impact of school poverty rate were also significant in explaining additional 2% variances in student academic score. The results of the models in this study supports several other research findings that points towards advantage of social economic status with regards to improvement in students academic achievement. In contrast (Snider & Bosker,1999) claimed that students characteristics account for more variances compared to schools.

Any improvement measures to upgrade education must focus on improving both school and students factors such as motivation and interest to learn. Eradicating poverty incidents in schools would improve students academic achievement substantially. This is exploratory study is not exhaustive and do not account for all variances at school level and students. More variables could be explored both at students and school level to improve accuracy of the model in future

7. Acknowledgement

This research was financially supported by the Fundamental Research Funds FRGS (Grant No. 600-RMI/ST/FRGS 5/3Fst28/2010) Malaysian Higher Institution.

8. References

Bryk, A.S. and Raudenbush S.W., (1994).Towards a more approriate conceptualization of research
On school effects:A three level linear model, American Journal of Education, 97(1), pp 65-108,
Peter F., Kloeckner N., Dalbert C.and Banerjee A. (2012). Belief in a just world, teacher justice, and student achievement: A multilevel study, Learning and individual differences 22,pp 55-63,