# Multilevel Modeling of Indonesian Student Achievement in Mathematics Based on TIMSS Data

Wiwit Pura Nurmayanti, Khairil Anwar Notodiputro, Indahwati

**Abstract** - Multilevel modeling can be used to analyze hierarchical data. The problem with hierarchical data occurs due to the individuals in the same group tend to be similar, thus between observations at a lower level may not be independent. This could be happened to TIMSS (Trends in International Mathematics and Science Study) data in which the students are nested within schools. In this paper we discussed the application of this model to TIMSS data and presented the results. Multilevel regression models were employed and in this research and it was shown that self-efficacy of students towards mathematics, students' attitudes toward mathematics, students' attitude towards teachers, students' attitude towards school, gender, the care of parents, frequency of homework, and violence/ bullying experienced by students, as well as the importance of mathematics for students have affected students' achievement in mathematics. Other factors at the school level such as the percentage of students who are less financially and school location were also significantly influences the achievement.

**Index Terms**— Multilevel modeling, hierarchical data, TIMSS, mathematics, self-efficacy, students' attitude, students' achievement.

## 1 Introduction

Multilevel modeling is a statistical technique to analyze data in a hierarchical structure. Hierarchical structure indicates that the analyzed data consists of several levels, in which the lower level is nested within a higher level (Hox, 2002). The reason is because the need for multilevel analysis in a hierarchical structure of individuals in the same group tends to be similar, so that between observations at a lower level are not independent. Multilevel model is part of a linear mixed models. This model can be applied to almost all fields such as health, marketing, social, economic, and education is no exception. The implementation of multilevel regression has belong recognized in the field of education because of the nature and relationships of the most of the data that exist in the field of education are hierarchical. The success of an educational system is basically affected by several things, such as the factor of students, parents, educational institutions, teachers, and environmental. As performed by Tantular (2009) who using multilevel regression, he found that someone's education (years) is affected by gender, parental education, status of residence (urban/rural), the number of schools in the residence (district), and the percentage of farmers. Other research in the field of education known as TIMSS (Trends in International Mathematics and Science Study) is conducted by an international research institutions engaged in educational evaluation, IEA (International Association for the Evaluation Educational Achievement).

TIMSS is an international assessment related to the knowledge of Mathematics and Science for students in fourth and eighth grade. Researches related to TIMSS data has been conducted by Riswan (2010) who performed the students math achievement grouping using logistics latent class analysis, while Widiastuti (2011) examined the factors that affect student achievement in mathematics using multilevel modeling. Widiastuti used TIMSS data of 2007 and the result showed that the model was not fit with $R^2$ at the student level and the school level that were still very small 0.59% and 1.57% respectively. So that, it needed improvement on the models with the addition of indicators that serve as explanatory variables at students level and the school level. The purpose of this study is to compare the mixed model using random intercept and random coefficient.

## 2 RESEARCH METHOD

### 2.1 Data

The study was based on survey data of TIMSS 2011 through the Center for Educational Assessment (Puspendik), the Research and Development Agency of the Ministry of National Education. Based on TIMSS data, the subjects were students in eighth grade in junior high school in the academic year of 2011. The observed variables in this study consisted of dependent variable, level-1 explanatory variables, and level-2 explanatory variables. Dependent variable was the student achievement, while level-1 explanatory variables (students) were gender (X1), self-efficacy of students towards mathematics (X2), students' attitudes toward mathematics (X3), students' attitudes towards school (X4), students' attitude towards teachers

- *Wiwit Pura Nurmayanti is currently pursuing masters degree program in applied statistics in Bogor Agricultural University, Indonesia, PH-+62877 6346 8152 E-mail: wiwit.adiwinata3@gmail.com*
- *Khairil Anwar Notodiputro is Lecturer, Departement of Statistics, Bogor Agricultural University, Bogor, Indonesia. E-mail: khairilnotodiputro@gmail.com*
- *Indahwati Lecturer, Departement of Statistics, Bogor Agricultural University, Bogor, Indonesia. E-mail: indah_stk@yahoo.com*

(X5), the importance of math for students (X6), the care of parents (X7), violence/ bullying experienced by students (X8), the number of books owned by the students (X9), and frequency of homework provided by the teacher at school (X10). Explanatory variables at level-2 (school) were the percentage of low level of economic of students (Z1) and the status of the school area (Z2).

### 2.2 Methods of data analysis

The data structure used was the student as an observation unit at level-1, and the school was the level-2. For the purposes of this analysis, then determined the following notations: $i$ = student ($i$ = 1, 2, ..., $nj$); $j$ = school ($j$ = 1, 2, ..., $J$), and $ij$ = $i$-th students in $j$-th school, and variables explanatory for students of $p$ = 1, 2, ..., $q$ and variables explanatory for school of $a$ = 1 , 2, ..., $b$. The notation was used in the analysis stage as follows:

1. Descriptive analysis of student and school data to observe the general overview
2. Modeling explanatory variables (level of students and schools) with the response variable (students achievement scores) (Hox, 2002)
   a. Random intercept model
      The intercept model is a model with intercept of random component, and the slope remains random. It is useful to determine the effect of the school on the model. In this model, variables at the school level were not included
      $$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \cdots + \beta_{10j}X_{10ij} + \varepsilon_{ij} \quad (1)$$
      with
      $$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + \gamma_{02}Z_{2j} + u_{0j}$$

      Note
      Yij a score of achievement at $i$-th student in the $j$-th school; $\beta_{0j}$ intercept; $\gamma_{0a}$ regression coefficient; $Zaj$ explanatory variables at school; $\beta_{pj}$ regression coefficient (slope); Xpij $p$-th explanatory variables at $i$-th students in $j$-th school; εij error at $i$-th students in $j$-th school. $\varepsilon_{ij} \sim N(0, \sigma_e^2)$
   b. Coefficient intercept model
      In random coefficient models considered the influence of the school. Likewise, the explanatory variables at the school level are also included as well as observing the interaction between the explanatory variables at level-1 (students) and level-2 (school), in the multilevel model with random coefficient, school influences is added at $p$-th variables.
      $$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \cdots + \beta_{10j}x_{10ij} + \varepsilon_{ij} \quad (2)$$
      with
      $$\beta_{0j} = \gamma_{00} + \gamma_{01}z_{1j} + \gamma_{02}z_{2j} + u_{0j}$$
      $$\beta_{pj} = \gamma_{p0} + \gamma_{p1}z_{1j} + \gamma_{p2}z_{2j} + u_{0j}$$
3. Selection of the best model
   The best model was determining by selecting the smallest value of Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC).
4. Estimation variance component of student achievement based on the model that has been obtained.

The variance of responses that can be explained by the explanatory variables in the regression model is called coefficient of determination. Coefficient of determination for each level could also be obtained in the multilevel model which was : (Bliese 2013) :

$$R_1^2 = 1 - \frac{\hat{\sigma}_{ep}^2}{\hat{\sigma}_{e0}^2} \quad (3)$$

$\hat{\sigma}_{ep}^2$ was the variance estimator of error at level-1 with $p$ explanatory variables, $\hat{\sigma}_{e0}^2$ was the variance estimator of error at level-1 without explanatory variables.

$$R_2^2 = 1 - \frac{\hat{\sigma}_{u0p}^2}{\hat{\sigma}_{u0}^2} \quad (4)$$

$\hat{\sigma}_{u0p}^2$ was the variance estimator of error at level-2 with $p$ explanatory variables, $\hat{\sigma}_{u0}^2$ was the variance estimator of error in the level-2 without explanatory variables.

## 3 Results and Discussion

Based on the results of the TIMSS survey in 2011, Indonesian student achievement in mathematics and science was still low when compared to other participating countries. Among 42 participant countries, Indonesia were 38th in math, and 40th in science. Students who participated in TIMSS in 2011 were 4567 students from 153 schools. Here is an overview of the achievements, student and school background:

### 3.1 Category of student achievements

Both mathematics scores data had been grouped by TIMSS based on Math International Benchmark into five categories:

- Very low (< 400)
- Low (400 - 474)
- Moderate (475 - 549)
- High (550 - 624)
- Advance (>624)

The information on Indonesian student achievements are presented in Fig 1:

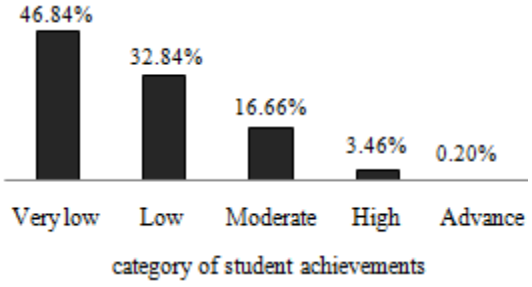
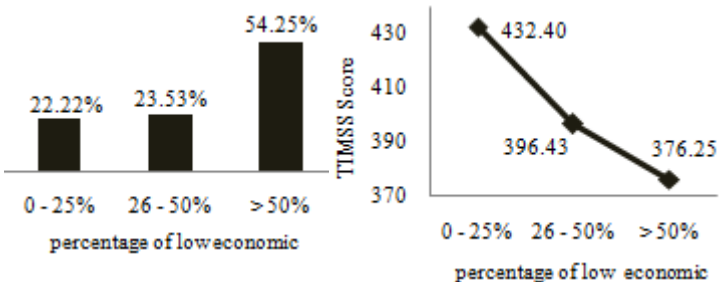



Fig 1 Percentage of students based on achievement categories

The data showed average achievement mathematics of Indonesian students were still in the very low category (46.84%). This means that Indonesian student achievement in mathematics was still relatively very low when compared to most other TIMSS participant countries. It is a problem, because mathematics is important to learn. According to Salim (2010), understanding of Indonesian students to math will determine the degree of their

readiness to learn a various knowledge and skills in higher education, and these subjects also always be a priority in all countries because they involve the understanding of technology and industry. Furthermore, analysis to determine the cause of the low level of achievement of students, especially in the TIMSS study, should be conducted.

### 3.2 Percentage of low level economic

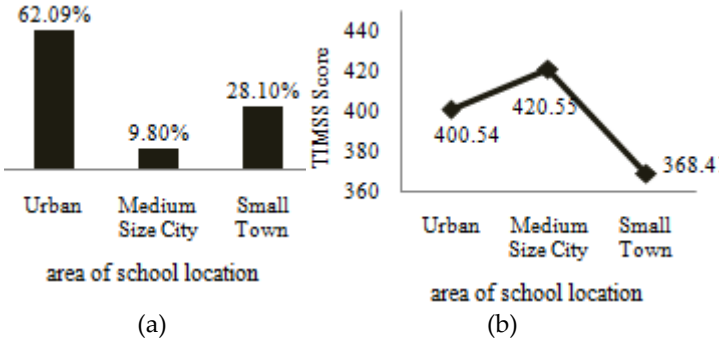



(a) (b)

Fig 2 (a) The percentage of schools based on the percentage of low level of economic, (b) The average of student achievement based on the percentage low level of economic

Fig 2 provides information that there was a tendency when the percentage of economic level of students in a school is above 25%, then the average achievement of student in this school is below the average Indonesian student achievement in general. In contrast, when the percentage of economic level of students in a school was below 25%, the average achievement of student is above average in general. This indicates that the economic problems have an impact on student achievement (maybe) indirectly, but in general, the economic condition of parents supports the students in their learning facilities, which in turn would facilitate and help the school to improve teaching and learning process.

### 3.3 Status of area




(a) (b)

Fig 3 (a) Percentage of schools based on status of the area, (b) The average of student achievement based on the status area

Fig 3 provides information that the schools in medium size city tend to had student achievement above the average of Indonesian students in general. In contrast, schools in urban areas had student achievement in below the average. In fact, urban area is developed or better area than medium size city when observed from the aspect of facilities, access, and the environment. Student achievement in schools of the small town tend to be far below the average and very low category, this may be due to the facilities, access, and environments that exist in the area of the small town is not as good and complete as in urban areas and medium size city, so there are a lot of limitations or constraints faced by students and teachers in the education process.

### 3.4 Multilevel Modeling

Intercept is the value of y when x = 0 as indicated by the regression line and the y-axis, and the slope indicates the slope of the regression line. When the regression analysis is applied to the group, it will get the regression lines which are equal to the number of groups. Parameter of intercept and the group could be different. So that, it's applied with a multilevel regression analysis with two models which are random intercept model and random coefficient model. Modeling the random intercept is done to determine whether there is any difference in the intercept of each school, and the formation of a random coefficient model is used to determine whether there is any difference in slope of each school. Here are the results of multilevel analysis model with random intercept and random coefficient:

TABLE 1
The Results of Analysis of the Multilevel Model of Math Achievement

| | Random Intercept | | | Random Coefficient (X3 X4 X5) | | |
|---|---|---|---|---|---|---|
| | Estimate | Std. Error | p-value | Estimate | Std. Error | p-value |
| (Intercept) | 336.51 | 18.28 | 0.00 | 345.28 | 18.14 | 0.00 |
| X1 (gender) | -3.20 | 1.60 | 0.05 | -3.37 | 1.60 | 0.04 |
| X2 (self-efficacy) | 16.37 | 2.34 | 0.00 | 15.75 | 2.33 | 0.00 |
| X3 (students attitudes toward math) | 34.21 | 2.46 | 0.00 | 33.25 | 2.75 | 0.00 |
| X4 (students attitude towards school) | -4.88 | 2.07 | 0.02 | -4.79 | 2.07 | 0.02 |
| X5 (students attitudes toward teachers) | -18.14 | 2.11 | 0.00 | -17.75 | 2.56 | 0.00 |
| X6 (the importance of math for students) | 4.04 | 1.91 | 0.03 | 3.75 | 1.91 | 0.05 |
| X7 (the care of parents) | 7.46 | 1.10 | 0.00 | 7.19 | 1.10 | 0.00 |
| X8 (violence/bullying experienced by students) | -5.96 | 1.17 | 0.00 | -5.99 | 1.17 | 0.00 |
| X9 (the number of books owned by the students) | - | - | - | - | - | - |
| X10 (frequency of homework) | -6.06 | 1.05 | 0.00 | -6.12 | 1.05 | 0.00 |
| Z1 (percentage of low economic 0 - 25%) | 45.17 | 11.21 | 0.00 | 42.07 | 10.92 | 0.00 |
| Z1 (percentage of low economic 26 – 50%) | 15.74 | 10.09 | 0.12 | 15.00 | 9.91 | 0.13 |
| Z2 (school location: Small Town) | -34.49 | 16.16 | 0.03 | -36.91 | 15.86 | 0.02 |
| Z2 (school location: Urban) | -4.96 | 14.63 | 0.74 | -8.77 | 14.32 | 0.54 |

Based on TABLE 1 shows that at a significance level of 5% both models random intercept and random coefficients of all the variables in level-1 (students) except X9 (the number of books owned by the students) and level 2 (school) have a significant effect on student mathematics achievement. The next step is choosing the best model to see the smallest AIC and BIC value.

TABLE 2
Selection of the best model

| value | Random Intercept | Random Coefficient |
|---|---|---|
| AIC | 49418.87 | 49425.09 |
| BIC | 49579.46 | 49527.86 |
| RMSE | 2458.60 | 2567.96 |
| Log.lik | -24684.43 | -24696.54 |

In TABLE 2 shows that the smallest AIC, RMSE and Log.lik value in the multilevel model with random intercept, thus it can be concluded multilevel model equations with random intercept is the best model compared to the random coefficients models. In addition, by using equations 3 and 4 obtained $R^2$ values for each level either at intercept random models and random coefficient. Here are the results of $R^2$ at every level in both models:

TABLE 3
Variance of random intercept models

| | Without the explanatory variables | Random Intercept | | Random Coefficients | |
|---|---|---|---|---|---|
| | Variance | variance | $R^2$ | variance | $R^2$ |
| Level-1 | 3022.04 | 2417.00 | 0.20 | 2590.00 | 0.14 |
| Level-2 | 2825.98 | 2515.24 | 0.11 | 2935.41 | -0.04 |

These results imply that the diversity of students' mathematics achievement can be explained by the explanatory variables at level-1 amounted to 20.00% greater than the random coefficient, and the rest is explained by other variables that have not been included in the modeling. While at level-2, the diversity of student achievement can be explained by the school factors amounted to 11.00% greater than the $R^2$ at random coefficient models.

Previous research about TIMSS data was ever done by Widiastuti (2011) and the result showed that the model is not yet fit with the $R^2$ at all levels were so small, they are the student level 0.59% and the school level 0.57%. So this research tried to fix the model on previous research by adding the indicators used as explanatory variables both at the student and school levels, and obtained the results that the model was already fit to the $R^2$ at all levels better than in previous studies that the student level 20.00% and the school level 11.00%.

## 4 Conclusion

The conclusion of this study is a random intercept models is the most suitable model to describe the influence of variables at level-1 on the mathematics achievement of students by taking into account the diversity among schools. Factors affecting the mathematics achievement of students at level 1 (students) are gender, self-efficacy of students' towards mathematics, students' attitudes toward mathematics, students' attitudes towards school, students' attitudes towards teachers, the importance of math for students, the care of parents, violence/ bullying experienced by students, and frequency of homework provided by the teacher at school, and at the level-2 (school) including the percentage of low level of economic of students and the status of the school area.

## REFERENCES

[1] Tantular B, "Penerapan Model Regresi Linear Multilevel pada Data Pendidikan dan Data Nilai Ujian,"thesis, Institut Pertanian Bogor, 2009.
[2] Riswan B, "*Analisis Logistik Kelas Laten (Pengelompokan Prestasi Matematika Siswa Indonesia Berdasarkan Hasil Survey TIMSS)*," thesis, Institut Pertanian Bogor, 2010.
[3] Widiastuti M, "*Kajian terhadap Faktor-faktor yang Mempengaruhi Pencapaian Siswa Bidang Matematika Menggunakan Pemodelan Multilevel*," thesis, Institut Pertanian Bogor, 2010.
[4] Hox J. 2002, *Multilevel Analysis Techniques and Applications*, New Jersey: Lawrence Erlbaum Associates, Inc.
[5] Bliese P, Multilevel Models in R (2.5). R Development Core Team, 2013.
[6] Salim, "*Determinants Of Learning Outcomes an Analysis Using Data The Trend In International Mathematics and Science Study (TIMSS) 2007*," Laporan Pusat Penilaian Pendidikan, Badan Penelitian dan Pengembangan Departemen Pendidikan Nasional, 2010.
[7] http://timss.bc.edu/.