A multilevel Analysis of Student and School Factors Influencing Primary Students' Science Achievement in TIMSS 2011 in East and South-East Asian Countries

Pongprapan Pongsophon

Science Education Program, Faculty of Education, Kasetsart University, Thailand 10900 Email: feduppp@ku.ac.th

Abstract

This study aims to examine and explained within- and between-school variance of science scores of grade 4 students in TIMSS 2011 from sixAsean and East Asian countries. The level-1 predictive variables were gender, parental involvement, perception on teacher's teaching practice and self-efficacy. The level-2 predictive variables included school size, learning resource, parent-school partnership, social environment, and principal's leadership. Multilevel effects were examined by Hierarchical Linear Modelling with HLM 7 software using TIMSS 2011 dataset. Regarding descriptive and one-way ANOVA statistics, most top performing countries had lower means in all level-1 predictive variables but higher in level-2 predictive variables of general and science learning resource and social environment. All countries had relatively higher means score on school-parent partnership. Korea and Thai had highest means in educational leadership. In terms of variance components, Japan and Korea had the lowest between-school variance. Gender and self-efficacy had effect on science achievement in all countries. Being a small school and social environment had negative and positive effect on school means respectively in most countries. Educational leadership had no effect whatsoever on the school means. These findings have implications for policy makers and practitioners on how to raise quality of science education in poor performing countries.

Keywords: Multilevel Analysis, Science Performance, TIMSS 2011, ASEAN, East Asia

1. Introduction

Trends in International Mathematics and Science Study or TIMSS is an international evaluation of mathematics and science achievement. Started in 1995, TIMSS has been conducted every 4 years to compare mathematics and science performanceof Grade 4 and Grade 8 students froma number ofeducational systems worldwide and examine the growth or recessionin the achievement over the cycles. The project was initiated, has been administered and organized by The International Association for the Evaluation of Educational Achievement (IEA). Apart from measuring student achievement, TIMSS collects data on learning contexts covering home, classroom, school and curriculum factors that may have effect on student performance. It is widely accepted that TIMSS results are a reliable indicator for a country'scurrent and future science and technology- based economy and quality of science and mathematics education. TIMSS's achievement test and contextual questionnaires aredesigned and developed by collaborative teams of content experts, science educators and educational assessment and evaluation experts who are the representatives from the participating countries. At national level, the test is administered by national country coordinator who is trained by TIMSS and PIRLS International Study Center, Lynch School of Education, Boston Collegeand it strictly follows the IEA's guideline. The test administration ismonitored by independent outside observers as a quality control mechanism. Examining the TIMSS results comparatively, in particular from outperforming countries, broadens our views on effective science education policies and practices that would help guide policy makers and practitioners to make an informed decision on science and mathematics education) Baker &LeTendre, 2005; Chepete, 2008; Lee, & Smith, 1993; Schreiber, (2002

This project was started narrowly since the author had had a concern on Thailand, his home country's plunging performance and later extended his concern to many other poor performing countries in Asia. Having participated since 1995, Thai performance has been falling below the international benchmark. The gender gap has been widened over cycles in that girls outperformed boys while top Asian countries like South Korea, Japan, Singapore, and Taiwan have been clustered in the top end in the ranking and could successfully close the gap between girls and boys. Respond to this alarming call, the author would like to examine and compare the variance in science scores within and between schools in all participating countries in East and South-East Asia (ASEAN) and develop a hypothetical model from intensive reviews and test if and how well it could explain such two level variances. The findings would explain why and how the top performing countries get to the top of the world and what lessons that the poor could learn from and take to improve their science education to make them stay competitive in the global economy.

This project used data from TIMSS 2011, the latest cycle when this project was conducted. In TIMSS 2011, there were more than six hundred thousand Grade 4 and Grade 8 students from fifty two and fifty five education systems participating respectively. The data were retrieved from IEA' s data repository website that was provided and maintained for public use to promote further advanced studies. Taking into account the structure of data in TIMSS in which smaller units were aggregated to a higher level unit, multilevel analysis was used. In this project, hierarchical linear modeling was developed and tested using the analytical technique proposed by RaudenbushandByrk (2002) with the assistance of HLM software.

2. Research proposes

1. To compare the means of selected predictive variables in the hypothetical model
2. To examine and explain within and between variances in science scores of the participating countries
3. To compare the tested models of the participating countries in terms of strength and weaknesses

3. Theoretical framework and the hypothetical model

This study is based on socio-cultural perspective of learning that are based on students' participation in social interactions and culturally organized activities. This participation is guided as Rogoff (1990) writes, "Children's cognitive development is an apprenticeship; it occurs through guided participation in social activity with companions who support and stretch children's understanding of and skill in using tools of culture" (p. vii). Learning is regarded as a cultural and social activity and is situated and affected by a complex system of factors. To fully understand the impact of each of these factors on students' science achievement, this study selected a set of variables and modelled them not in isolation but in relation to others.

There have been extensive studies on the factors influencing student achievements(Damrongpanit&Ruengtrakul, 2013). To guide predictive variable selection, the Rand model (Oakes, 1986; Shavelson, McDonnell, & Oakes, 1989) was considered. The Rand model provides a model of a general system that comprises three components; input, process and output. Applied to education, the input variables are, for example, teacher quality, students' socio-economics status, learning resources. The process variables are what and how it is taught. The output variables are the concretely measurable outcomes of schooling such as academic achievement, attitudes, required skills and content, and reduction of youth unemployment. By nature, these factors are organized into different levels — from macro to intermediate and from intermediate to micro levels. The economic status of a country, region, or city serves as a proxy for a system's available resources and is located at the macro level of an education system. Factors associated with school and teachers are considered to operate at the intermediate level. Factors such as student gender, ability, familial support and attitudes toward school at the micro level.Taking this into consideration, a hierarchical linear modelling is used in this study. The units of analysis at a lower level were individual students who were nested within a school—more contextual/aggregate units which were the units of analysis at a higher level.

Based on the availability of variables in the original TIMSS contextual questionnaires, two set of predictive variables at student and school level were included in the model. These were found to have significant effect on science achievement in literature (Liu, Wu, &Zumbo, 2006). The predictive variables at student level were gender (Female), parental involvement in child's education (ParInv), student's perception of their teacher' teaching practice (Teach), and student's self-efficacy (SelfEff). Those at school level include school size (SmallSch), general resource (GenRes), science resource (SciRes), social environment (SocEnt), school-parent partnership (SchPar), and principal's educational leadership (Lead). The hypothetical model of multilevel factors influencing science achievement in TIMSS 2011 is illustrated in figure 1.

Figure 1: A hypothetical model of multilevel factors influencing science achievement
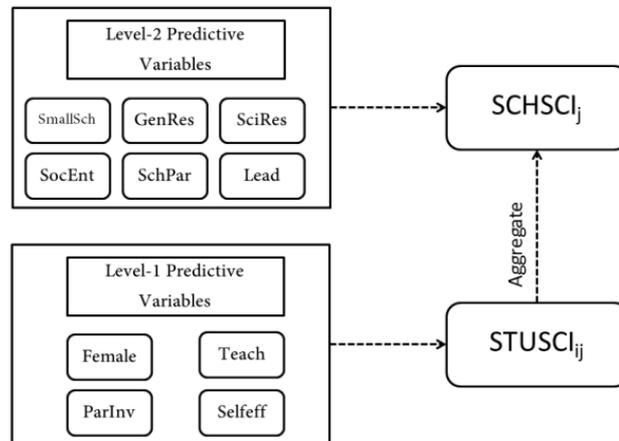
3. Methodology

3.1 Sample

The samples of grade 4 students and their schools of each country were randomly selected using stratified two-stage cluster sample design (Table 1). First, the schools were sampled with probabilities proportional to their size from the list of all schools in the population that contain eligible students. The schools in this list may be stratifiedaccording to important demographic variables. The second sampling stage consists of the selection of one or more intact class from the target grade of each participating school.

Table 1: Samples of primary students and schools by countries (before data screening)

	East Asia			ASEAN		
	Hongkong	Japan	S. Korea	Taiwan	Singapore	Thailand
Students	3,957	4,411	4,334	4,284	6,368	4,448
Schools	136	149	150	150	176	168

3.2 Instruments

TIMSS 2011 uses a matrix-sampling approach that involves packaging the entire assessment pool of mathematics and science items into a set of 14 student achievement booklets, with each student completing just one booklet. The assessment time for each booklet was limited to 90 minutes; 72 minutes for science and mathematics achievement tests and 30 minutes for a student questionnaire. TIMSS achievement test integrated content knowledge and cognitive skills. The content covers biological, physical and earth science. The cognitive skills cover knowing, applying and reasoning. The achievement test consists of four item blocks; two for science and the other two for mathematics. The students took 18 minutes to complete each block. In each block, about half of the items were multiple-choice and the other half was constructed-response items.The items were administered in 28 item blocks, 14 blocks for science and 14 blocks for mathematics. Student booklets were assembled from various combinations of these item blocks. Booklet design is illustrated in Table 2.Each item appears in two booklets, providing a mechanism for linking together the student responses from the various booklets. TIMSS uses item-response theory scaling methods to assemble a comprehensive picture of the achievement of the entire student population from the combined responses of individual students to the booklets that they are assigned. Five plausible values forthe individual student's total score were estimated. These values were included in HLM analysis.

Table 2: Booklet design for a student booklet in TIMSS 2011

Booklets	Assessment Blocks for Achievement Test				questionnaire
	Part I		Part II		
Booklet 1	M01	M02	S01	S02	✓
Booklet 2	S02	S03	M02	M03	✓
Booklet 3	M03	M04	S03	S04	✓
Booklet…	…	…	…	…	✓
Booklet 13	M13	M14	S13	S14	✓
Booklet 14	S14	S01	M14	M01	✓

3.3 Data screening, cleaning, and creating new variables

The secondary data were retrieved from IEA's repository website. There were two files downloaded; student and school files both in sav.file format. Only items of predictive variables in the hypothetical model were selected; others deleted. Except Female and SmallSch, the predictive variables were continuous. The files were screened for any missing values. The cases with missing values of continuous variables were treated using Expectation-Maximization algorithm. Cases with unsuccessful imputation of missing values and those with missing value of discrete variable were deleted before running multilevel analysis in HLM. For the continuous variables, their values were the means of underlying items; summated scaling. For particular items, recoding of data is necessary to prevent misinterpretation of the final results. These items used small numbers to represent large values. The recoding is also used for making dummy variables for Female and SmallSch; designating the value of 1 for female and 0 for male; 1 for a small school and 0 for other school sizes. Data screening,treatment, and manipulation were done in SPSS software. The predictive variables, samples oftheir underlying items, and Conbrach's coefficients were summarized in Table 3.

Table 3a:Level-1 predictive variables, number of items, underlying items, reliability coefficients.

Predictive variables	Cronbach's alpha	No. of Items	Sample of the statements
Student Level			
Female	-	-	-
Parental Involvement (ParInv)	0.83	4	How often do the following things happen at home? - My parents ask me what I am learning in school - I talk about my schoolwork with my parents
Student Perception of Teacher' Teaching Behavior (Teach)	0.76	5	How much do you agree with these statements about your science lessons? - I know what my teacher expects me to do - I am interested in what my teacher says
Self-Efficacy(SelfEff)	0.86	9	How much do you agree with these statements about science? - I usually do well in science - Science is harder for me than any other subject

Table 3a: Level-2 predictive variables, number of items, underlying items, reliability coefficients.

School Level			
School Size (SmallSch)	-	-	-
General Learning Resource(GenRes)	0.74	6	How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following? - Instructional materials (e.g., textbooks) - Supplies (e.g., papers, pencils)
Science Learning Resource (GenSci)	0.77	7	How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following? - Computer software for science instruction - Science equipment and materials
Social Environment(SocEnt)	0.72	11	To what degree is each of the following a problem among students in your school? - Classroom disturbance - Intimidation or verbal abuse among students
School-Parent Partnership (SchPar)	0.73	13	How often does your school do the following for parents in general? - Inform parents about the overall academic achievement of the school - Discuss parents' concerns or wishes about the school's organization
Principal's Educational Leadership(Lead)	0.81	13	During the past year, approximately how much time have you spent on the following school leadership activities in your role as a school principal? - Promoting the school's educational vision or goals - Initiating educational projects or improvements

3.4 Data analysis

3.4.1 Descriptive statistics and one-way ANOVA

To summarize each set of the data, means and standard deviation of all predictive variables at both levels were calculated. Next, one-way ANOVA was utilized to determine whether mean scores of the predictive variables between countries differed. If verified that not all mean scores were equal, multiple comparison tests were run to analyze the exact relationships between them.

2.4.2 Multilevel analysis

The outcome variable and all predictive variables at both levels except Female and SmallSch were standardized using SPSS prior to the HLM analysis. HLM output would then give standardized coefficients as results so the regression coefficients of predictive variables can be compared with one another in terms of relative strength (Dallal, 2012).MDM files were prepared in HLM. These were read and used for modelling in HLM. Raudenbush and Bryk (2002) proposed a technique for multilevel analysis in which four models were run in order. The models were increased in complexity. The results of the former models determine model adjustment and guide the decision on the creation of the latter. The first model was a one-way random effect ANOVA, the simplest possible random effect linear model. This model examined within-and between -school variances when no predictive values added in the model where $r_{ij} \sim N(0,\sigma^2)$ and $u_{0j} \sim N(0, \tau^2)$. Interclass correlation (ICC: $\hat{\rho}$), proportions and percentages of variance of both levels can be calculated from the variation at level-2(τ^2) and the variation at level-1 (σ^2) estimated by this model. The intraclass correlation coefficient represents the proportion of the variance in outcome between the schools. As such, this model serves as baseline for comparison with subsequent, more complex models. The ICC can be computed in the following equation.

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2}$$

The second and the third models are named regression with means-as-outcomes and a random-coefficient model respectively. In the second model, only level-2 predictive variables are added to the model and were tested their effects on the school means. The model also helps us examine if and how much the addition canreduce level-2 variance. The third model works similarly to the second model but only adds level-1 predictive variables to the model. The third model examines the effect of the level-1 predictive variable and examines if and how much level-1 variance is reduced and explained. The equations of variance reduction of the second and third models are shown in Table 4.

Table 4: Equations for variance reduction in models 2 and 3

Variance reduction	
At Level-2	At Level-1
$\dfrac{\tau^2_{model1} - \tau^2_{model2}}{\tau^2_{model1}}$	$\dfrac{\sigma^2_{model1} - \sigma^2_{model3}}{\sigma^2_{model1}}$

The forth model is called an intercepts- and slopes-as-outcomes model. It is the hypothetical model of the multilevel analysis. Both level-1 and level-2 predictive variables are added to the model and tested for their effect on the outcome variable. In this model, the reduction and implied explanation of variance at both levels can be computed. The equations of all model are summarized in Table 5.

Table 5: Multilevel models and their multi-equation and mixed equation formats

Model1: Unconditional means model
Level-1 Model
$STUSCI_j = \beta_{0j} + r_{ij}$
Level-2 Model
$\beta_{0j} = \gamma_{00} + u_{0j}$
Model2: regression with means-as-outcomes
Level-1 Model
$STUSCI_j = \beta_{0j} + r_{ij}$
Level-2 Model
$\beta_{0j} = \gamma_{00} + \gamma_{01}*(SMALLSCH_j) + \gamma_{02}*(GENRES_j) + \gamma_{03}*(SCIRES_j) + \gamma_{04}*(SCHPAR_j) + \gamma_{05}*(SOCENT_j)$ $+ \gamma_{06}*(LEAD_j) + u_{0j}$
Model 3: random-coefficient model
Level-1 Model
$STUSCI_j = \beta_{0j} + \beta_{1j}*(FEMALE_{ij}) + \beta_{2j}*(PARINV_{ij}) + \beta_{3j}*(TEACH_{ij}) + \beta_{4j}*(SELFEFF_{ij}) + r_{ij}$
Level-2 Model
$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10}$
$\beta_{2j} = \gamma_{20}$
$\beta_{3j} = \gamma_{30}$
$\beta_{4j} = \gamma_{40}$
Model 4: intercepts- and slopes-as-outcomes model
Level-1 Model
$STUSCI_j = \beta_{0j} + \beta_{1j}*(FEMALE_{ij}) + \beta_{2j}*(PARINV_{ij}) + \beta_{3j}*(TEACH_{ij}) + \beta_{4j}*(SELFEFF_{ij}) + r_{ij}$
Level-2 Model
$\beta_{0j} = \gamma_{00} + \gamma_{01}*(SMALLSCH_j) + \gamma_{02}*(GENRES_j) + \gamma_{03}*(SCIRES_j) + \gamma_{04}*(SCHPAR_j) + \gamma_{05}*(SOCENT_j)$ $+ \gamma_{06}*(LEAD_j) + u_{0j}$
$\beta_{1j} = \gamma_{10}$
$\beta_{2j} = \gamma_{20}$
$\beta_{3j} = \gamma_{30}$
$\beta_{4j} = \gamma_{40}$

4. Research Findings

Comparing country means of science achievement at Grade 4 level (Table 6), South Korea had the highest score followed by Singapore. Thailand had the lowest and was below the international benchmark of 500 scores. By gender, boys outperformed girls by 4-7 scores in all countries except Thailand. Singapore had narrowest gap in scores between boys and girls. Thailand, reversely, girls did better than boys by 9 scores, the widest gap compared to other countries. Regarding performance on specific areas of science, South Korea had the highest score followed by Singapore in Biological Science. Japan and Singapore did best in Physical Science followed by South Korea by just one score. South Korea had highest score in Earth Science; leaving the second place Taiwan by 50 scores. In the three areas of science, Thai primary students did best in Biological Science; they got 20 scores or so lower in the other areas.

Table 6: Grade 4 students' science achievement by gender and areas of science

	Mean Scores					
	HKG	JPN	KOR	TWN	SGP	THAI
All	535	559	587	552	583	472
Girls	532	556	583	548	581	476
Boys	538	561	590	555	585	467
Areas of Science						
Biological science	524	540	571	538	597	480
Physical science	539	598	597	569	598	462
Earth science	548	551	603	553	541	460

4.1 Descriptive statistics

Descriptive statistics showed that (Table 7), for level-1 predictive variables, primary students in Japan and South Korea had lowparental involvement in their education, lowperception on their teachers' teaching practice, and lowself-efficacy compared with their counterparts in other countries. Thai students had their parents involved in their education the most and had the highest score onthe perception of their teachers' teaching practice. Taiwanese students had highest self-efficacy. Regarding level-2 predictive variables, most top performing countries; Japan, South Korea, and Singapore, were least suffered from the shortages of general and science learning resources in school. South Korea and Singapore, in addition, had more pleasant social environment in schools. School-parent partnership was generally high for all countries. Principals in Thailand and South Korea had higher educational leadership than those of other countries.

Table 7: Descriptive statistics of two level predictive variables

Level 1	East Asia								ASEAN			
	Hong Kong		Japan		Korea		Taiwan		Singapore		Thailand	
# of students	3,957		4,411		4,334		4,284		6,368		4,448	
Predictive variables	freq.	%	freq.	%	freq.	%	freq.	%	freq.	%	freq.	%
Female	1,839	46.50	2,177	49.35	2,085	48.11	2,029	47.36	3,135	49.23	2,211	49.70
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
ParInv	2.85	0.869	2.66	0.81	2.81	0.81	3.04	0.87	3.11	0.80	3.14	0.71
Teach	3.16	0.69	2.74	0.64	2.92	0.61	3.27	0.64	3.27	0.62	3.40	0.49
Selfeff	2.85	0.70	2.82	0.59	2.76	0.60	3.21	0.66	2.84	0.72	2.88	0.58
Level 2												
# of schools	136		149		150		150		176		168	
Predictive variables	freq.	%	freq.	%	freq.	%	freq.	%	freq.	%	freq.	%
SmallSch	24	17.65	74	49.66	17	11.33	25	16.67	1	0.57	89	52.98
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
GenRes	1.87	0.50	3.15	0.68	3.57	0.56	2.27	0.86	3.11	0.89	2.31	0.83
SciRes	2.27	0.46	2.75	0.68	3.36	0.68	2.21	0.79	2.90	0.81	1.83	0.76
SocEnt	3.00	0.40	2.86	0.42	3.24	0.43	3.22	0.47	3.16	0.39	2.68	0.40
SchPar	3.73	0.31	3.56	0.46	3.60	0.54	3.72	0.32	3.60	0.35	3.46	0.51
Lead	2.39	0.26	2.21	0.31	2.79	0.30	2.50	0.32	2.54	0.30	2.69	0.27

4.2 ANOVA results

A one-way between groups analysis of variance was conducted to compare means of predictive variables at both levels of all countries under the study. There was a statistically significant difference at the p <0.0001 level in all level-1 predictivevariable scores; ParInv, F (5, 27580) = 254.12, p <0.0001; Teach, F (5, 27527) = 732.47, p < 0.0001; and Self-Eff, F (5, 27516) = 268.60, p < 0.0001.The actual difference in mean scores between groups was quite small in ParInv and Self-Eff and large in Teach according to Cohen (1988). The effect size of the countries of the predictive variables, calculated using eta squared (η^2), was 0.04, 0.04, and 0.11 respectively. Post-hoc comparisons using the Scheffetest indicated that the mean score of all variables in most countries differ. Regarding level-2 predictive variables, the difference in means of all variables in all countries was statistically significant; GenRes, F (5, 923) = 115.44, p < 0.0001, η^2 = 0.38); SciRes, F (5, 923) = 94.47, p< 0.0001, η^2= 0.33; SchPar, F (5, 923) = 46.50, p < 0.001, η^2= 0.20; SocEnt, F (5, 923) = 8.60, p < 0.0001, η^2= 0.04; and Lead, F (5, 923) = 77.116, p < 0.0001, η^2= 0.29. The effect sizes of the countries on level-2 variables are very large except SocEnt that is quite small. Post-hoc comparisons using the Scheffetest indicated that the mean score of these variables for most countries differ.

4.3 HLM results

The analysis of variance components in the Model 1 of HLM analysis showed that at Grade 4, Japan and South Korea had the least variation in science achievement between schools (6%) while Thailand and Hong Kong are the two countries that varies the most in science achievement at school level by 37 and 38 percent respectively (Table 8).

Table 8: The estimation of between- and within-school variances by countries in Model 1

Countries	Between School(u_0)		Within-School (r)		ICC	Percentage of within-school variance	Percentage of between-school variance
	Variance component	SD	*Variance component*	SD			
Hongkong	0.44***	0.66	0.73***	0.86	0.38	62	38
Japan	0.07***	0.26	0.98***	0.99	0.06	94	6
S. Korea	0.06***	0.25	0.94***	0.97	0.06	94	6
Taiwan	0.17***	0.41	0.96***	0.97	0.15	85	15
Singapore	0.25***	0.5	0.75***	0.87	0.25	75	25
Thailand	0.34***	0.58	0.57***	0.76	0.37	53	37

***p <0.001

The amount of two level variances provides the baseline for subsequent analyses. To explain within- and between-school variances, the model 2, 3, and 4 were run; adding one level or both levels at a time in multilevel model to examine the reduction in the variance. Due to the page limit in this publication, only the results of testing hypothetical model in model 4 will be presented. The effect of a particular level-1predictive variables on the individual student's science achievement varied across countries (Table). Female had no significant effect in all countries except Hong Kong. Parental involvement showed mixed results. In Hong Kong, parent involvement negatively related to science scores while in Taiwan, it had positive effect. Others had no effects. Perception of teacher's teaching practice had effect on science achievement in Singapore and Thailand. The relation was significantly positive.Self-Eff was the only predictive variable that had effect in all countries. Moreover, It is the strongest and positive predictor. Regarding level-2 variables, their effect varies across countries. SmallSch had effect on school's science achievement in all countries except in Singapore that had no effect. Consensually, among level-2 variables, being a small school has strongest effect; it effects was negative meaning that small schools did more poorly than schools of other sizes. GenRes and SciRes had effect only in South Korea in negative and positive

way respectively. SchPar had positive effect in Taiwan and Singapore and SocEnt had positive effect in Hong Kong and Thailand. Lead, interestingly, had no effect on school achievement in science in any countries.

Table4.8: The robust estimation of fixed effectby countries (Model 4; hypothetical model)

Fixed effects	East Asia								ASEAN			
	Hong Kong		Japan		S. Korea		Taiwan		Singapore		Thailand	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept, γ_{00}	0.65***	0.06	0.96***	0.04	1.33***	0.04	0.87***	0.05	1.28***	0.04	-0.19***	0.10
Level-1 predictors												
Female, γ_{10}	-0.12**	0.04	-0.03	0.06	-0.06	0.06	-0.11	0.05	-0.04	0.03	0.04	0.05
ParInv, γ_{20}	-0.06**	0.02	0.02	0.02	0.02	0.02	0.12***	0.03	0.003	0.01	-0.01	0.02
Teach, γ_{30}	0.02	0.02	-0.04	0.03	-0.06	0.04	-0.04	0.03	0.04**	0.01	0.13***	0.02
Selfeff, γ_{40}	0.20***	0.02	0.26***	0.03	0.37***	0.04	0.36***	0.03	0.26***	0.01	0.15***	0.02
Level-2 predictors												
SmallSch, γ_{01}	-0.41*	0.17	-0.25***	0.05	-0.30***	0.07	-0.42***	0.11	-0.29	0.52	-0.28*	0.28
GenRes, γ_{02}	-0.03	0.08	-0.002	0.05	-0.08*	0.03	0.11	0.09	-0.02	0.05	0.05	0.05
SciRes, γ_{03}	0.02	0.06	-0.01	0.05	0.08*	0.03	-0.07	0.11	-0.01	0.05	0.04	0.04
SchPar, γ_{04}	-0.09	0.05	0.005	0.03	-0.02	0.04	0.09*	0.04	0.08*	0.04	-0.04	-0.04
SocEnt, γ_{05}	0.18*	0.09	0.01	0.04	0.03	0.03	0.07	0.08	0.06	0.04	0.12*	0.12
Lead, γ_{06}	-0.17	0.10	0.04	0.03	-0.006	0.03	-0.04	0.05	-0.07	0.04	-0.07	-0.07

* p< 0.05, **p<0.01, ***p <0.001
SE = Standard Error, Coeff. = Coefficien

Table4.8: The robust estimation of random effectby countries (Model 4; hypothetical model)

Random effects	East Asia								ASEAN			
	Hong Kong		Japan		South Korean		Taiwan		Singapore		Thailand	
	Var. comp.	SD	Var. comp.	SD	Var. comp.	SD	Var. comp.	SD	Var. comp.	SD	Var. comp.	SD
Between-school, u_0	0.32***	0.56	0.05***	0.23	0.04***	0.20	0.11***	0.33	0.24***	0.49	0.30***	0.55
Within-school, r	0.69***	0.82	0.92***	0.96	0.83***	0.90	0.82***	0.91	0.67***	0.82	0.53***	0.73
Level 1 variance explained (%)	5		6		11		14		11		7	
Level 2 variance explained (%)	27		28		33		35		4		12	

* p< 0.05, **p<0.01, ***p <0.001

5. Conclusion

5.1 Discussion

It is noticeable that the between-school variance in Japan and Korea was the smallest (6%) while that of Hong Kong (38%) and Thailand (37%) was much higher. This confirms the previous study of Kaya and Rice (1990) that between-school variance of science achievement in Japan in TIMSS 2003 was very small (4%) in contrast with Singapore that between-school variance was higher than 50 percent. Kay and Rice explained that such small between-school variance in Japan was due to its homogeneity in school systems that rooted in the society. Japan is a relatively homogenous country in terms of migration and ethnicity. In Japan, most students attend public schools through the lower secondary level; diversification in schooling starts at high school. Less than 1% of the primary schools are private. Primary school classes are also large. On the contrary, Singaporean schools are more diverse and westernized due to the racial make-up and the policy of decentralization in education. Influenced by Goh ChokTok's "Thinking Schools, Learning Nations" policy since 1997, Singaporean schools have been given autonomy in deciding their own curriculum and materials and student-ability driven. For example, to cater intellectually gifted students to their top potential, Singapore Ministry of Education launched the Gifted Education Program (GEP). Currently, there are nine primary schools offering this program.

This study found that the between-school variance was explained more than the within-school variance. This variance reduced and explained in most countries was less than 10 percent. This means that the hypothetical model must have ignored some key predictive variables that had been found having strong effect on science achievement such as student's socio-economic status, the time spent on doing homework, levels of literacy. Some of these variables had been initially considered by the author but later were excluded from the model because there was no or limited information about the variables in the original questionnaire.

When looked at the effect of individual predictive variable, only self-efficacy and being a small school had effect on school science achievement in most countries. Others showed mixed results. Gender generally has no effect. This confirms many previous studies that had found that there was no or narrow gender gap in science

performance at primary level (Grenfiled, 1995; Kahle, 2004). However, the gap was later widened in the secondary level. This is explainable in lenses of social psychology that secondary students like to work with their same-sex peers. Boys like to take a risk, get challenged, lead the group, and express their ideas in classroom. That is why boys outperform girls in previous TIMSS studies at secondary level. Boys are more recognized and paid more attention by their teacher (Dimitrov, 1999). This scenario is quite common in Asian context. The small gender gap in TIMSS 2011 at primary level can be pointed to successful policies in closing gender gap in education in some Asian countries. Korea, for example, has ensured gender equity in education and career. In 2012, Koreans celebrated their first female president-elect, Park Geun-hye, an engineeringgraduate. This is significant in the history of Korea and reflects the shifting normin the society. Founded in 2001, Korea Ministry of Gender Equality and Familywas established to promote gender equity for sustainable development. Regarded as equal humanpower, there were a number of related acts launched including Equal Employment Act in 1995, Women's Development Actin 1995, and five-year Comprehensive Plan for the Development of Women Resources and the Female Employment Expansion Measures in 2006.

This study found that parental involvement had mixed effect on science achievement; negative in Hong Kong and positive in Taiwan. The positive effect confirms the studies of Russell (1997) and Kaya and Rice (2010). Parental involvement means mental and financial support for more and better learning opportunity for the children but too much involvement can cause pressure, stress, and negative attitude toward learning science (Grolnick, Ryan, &Deci, 1991). Russell pointed to the big difference in the parental investment for their children in tutoring between the United States and top Asian countries. About 70 percent of parents in East Asia had paid for extra-school tutoring, almost three times higher than American parents. East Asian parents had higher expectation for their children. They thought this expense would guarantee better future for their loved ones in the increasing competitive job market.

The finding on the self-efficacy as the strongest and positive predictor on science achievement in all countries confirms the findings of Lieu, Wu and Zumbo (2006) and Elliot and Church(1997). Students with low self-efficacy toward a task are more likely to avoid it, while those with high self-efficacy are not only more likely to attempt the task, but they also will work harder and persist longer in the face of difficulties. Studentswith low self-efficacy don't have expectation and motivation in learning. They do not believe they have sufficient knowledge and skills to do well so they don't step out.

The finding on the negative effect of being a small school on school science achievement in most countries adds weight to some previous studies (Fowler & Walberg, 1991). In literature, the effect of being a small school on achievement varied; no effect, positive (Sander, 1993; Schreiber, 2002; Kantabutra& Tang, 2006) and negative effect. In western context, a small school has a number of advantages. A small school invites strong support from parents and community members as well as closer working relationships among the school staff. Teachers are more apt to know their students as individuals. The students are generally given more and equal learning opportunity in classroom. They interact more frequently and informally with the teacher and with each other. However, small school in some contexts, Thailand as an example, suffered tremendously from limited budget and shortages of science teachers and learning resources (Lieu, Wu &Zumbo, 2006).

5.2 Implication

We can learn how to raise science achievement and promote educational equity by examining models of science education in successful countries. Singapore and Korea have put much emphasis on inquiry based learning in science classroom. Both male and female students have given equal opportunity and are expected to work collaboratively and contribute to the community of learning; posing a question, designing and conducting scientific investigation, collecting, analyzing and interpreting data, drawing conclusion and applying scientific knowledge to solve real world problem. The teachers play vital role in establishingequitable classroom environment. They should be a facilitator providing thought-provoking yet attainable challenges, moral support and scaffolding for all students.

Regardless of gender, students grow in self-efficacy as they experience success in learning, just as they lose it in the face of repeated failure. Today, there are more world-recognized female Asian scientists;this could be brought in science classroom toinspire and motivate the girls. This could boost up their self-efficacy and positive attitude toward science. Love and understanding areindispensable in parental involvement in their child's learning. They should not over-schedule their child and listen to what their child loves and is good at and back it up. The parents should be a role model for learning. They might even establish a joint study time; for example, setting aside time to read together. The parents should lead children to read, talk about what they are reading, discuss the main ideas, new words and concepts, and ask about their favorite section. Regarding small school issue, rather than closing, as the case in Thailand, the government should ensure that children in remote areas have access to quality education and enroll in a school nearby their homes. Teacher capacity is built up whereas class is reduced in size and fully equipped with ICT. The government should engage communities in local decision-making on school management and governance.

6. Acknowledgements

References

Baker, D., and LeTendre, G. (2005).*National Differences, Global Similarities: World Culture and the Future of Schooling*. Stanford CA: Stanford University Press.

Chepete, P. (2008). *Modeling of the Factors Affecting Mathematical Achievement of Form 1 Students in Botswana based on the 2003 Trends in International Mathematics and Science Study*. (Unpublished Doctoral dissertation.), Indiana University, Indiana, USA

Cohen, J. (1988). *Statistical power analysis for the behavioral* sciences (2nded.). New Jersey: Lawrence Erlbaum.

Dallal, E. G. (2012).*Which predictors are more important?*Retrieved from www. jerrydallal.com/LHSP/importnt.htm

Damrongpanit, S., &Ruengtrakul, A. (2013). Effects of learning style, school sector and school size on academic achievements of ninth-grade students in northeast area. *Journal of Research Methodology*, 26(1), 1-42.

Dimitrov DM. 1999. Gender differences in science achievement: Differential effect of ability, response format, and strands of learning outcomes. *School science and mathematics*, 99(8), 445-450.

Elliot, A. J., & Church, M. (1997).A hierarchical model of approach and avoidance achievement motivation.*Journal of Personality and Social Psychology, 72*, 218–232.

Kahle, J. B. (2004). Will girls be left behind? Gender differences and accountability.*Journal of Research in Science Teaching, 41*(10), 961–969.

Kaya, S. &Ric.D.C., 2010. Multilevel Effects of Student and Classroom Factors on Elementrary Science Achievement in Five Countries. *International Journal of Science Education*. 32(10), 1337-1363

Fowler, W. J., Jr., & Walberg, H. J. (1991).School size, characteristics and outcomes.*Educational Evaluation and Policy Analysis, 13*, 189–202.

Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L., &Akey, K. L. (2004).Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation.*Contemporary Educational Psychology*, 29, 462–482.

Grolnick, W. S., Ryan, R. M., &Deci, E. L. (1991). Inner resources for school achievement: Motivational mediators of children's perceptions of their parents. *Journal of Educational Psychology, 83*, 508–517.

Kantabutra, S. & Tang, J. C. S. 2006. Urban-Rural and Size Effects on School Efficiency: The Case of Northern Thailand. *Leadership and Policy in Schools*, 5(4): 355-377.

Lee, V., & Smith, J. B. (1993).Effects of school restructuring on the achievement and engagement of middle-grade students.*Sociology of Education*, 66, 164–187.

Liu, Y., Wu, A.D. &Zumbo, B.D. (2006). The Relation between outside of school factors and mathematics achievement: a cross-country study among the U.S. and five top-performing Asian Countries.*Journal of

Educational Research & Policy Studies, 6(1), 1-35.

Oakes, J. (1986). *Educational indicators: A guide for policymakers*. Wisconsin-Madison: Center for Policy Research in Education.

Raudenbush, S. W., &Bryk, A. S. (2002).*Hierarchical linear models: Application and data analysis methods* (2nd Ed.). California: Sage Publications, Ltd.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Russell, N. U. (1997). Lessons from Japanese cram schools. In W. K. Cummings & P. Altbach (Eds.), *The challenge of Eastern-Asian education: Lessons for America* (pp. 153–170). Albany, NY: State University of New York Press.

Sander, W. (1993). Expenditures and student achievement in Illinois: New evidence. *Journal of Public Economics, 52*(3), 403-416.

Schreiber, J. B. (2002). Institutional and student factors and their influence on advanced mathematics achievement. *Journal of Educational Research*, *95*, 274–286.