ANALYSIS AND MODELING OF EDUCATION PARTICIPATION INDEX (EPI) IN INDONESIA FROM 2003-2008

Mustofa Usman
Department of Mathematics, Faculty of Sciences,
Universitas Lampung, Indonesia

Hendarman
Head of the Central of Educational Research, Ministry of Education Republic of Indonesia, Jakarta

Jamal I.D and Faiz A.M.Elfaki
Departments of Sciences, Faculty of Engineering, IIUM, Malaysia

ABSTRACT

The aims of the research is to reveal the characteristics of the Education Participation Index (EPI) in Indonesia based on the level of students' age (7-12, 13-15, 16-18, and 19-24) which shows the participation index of the citizens at Elementary School, Junior High School, Senior High School and University. The data was taken from Central Bureau Statistics of Indonesia (BPS) from the year 2003 to 2008. The data is analyzed to see the difference between the level of ages at difference regions and difference years. And the data was analyzed by using analysis nested design. The second analysis is to find the EPI model for each regions and years. The modeling is used the multiple linear regression with dummy variable for the regions and years.

Keywords: *Education participation index (EPI), nested design, multiple regressions, and dummy variables.*

1 INTRODUCTION

In many countries, the statistics of Education Participation Index(EPI) are very important statistics for planning the education development in the countries, to plan the job market in the future, and to plan the budgeting the education for their people. The statistics of Education Participation Index is the ratio of the total number of citizens at the level of certain range of age with the total number of the pupils who are in formal education. There are several research studies which suggest that increased compulsory or voluntary participation in education has many benefits for individuals; such studies appear to show that ‘children who would otherwise leave school early are, in fact, better off if they stay, or that society benefits collectively because a higher level of educational attainment promotes good citizenship and economic development’ (Oreopoulos, 2006a). The youth unemployment is being raised by falling participation in education (Mark Corney, 2009). There are many researches shows that the access to postsecondary education is a central policy issue in modern societies. Increased participation in postsecondary education is an important social goal as it is a crucial determinant of the economic success of an individual as well as of society as a whole. Because of the positive benefits of postsecondary education, equitable access to postsecondary education for individuals from all backgrounds can ensure a higher level of social mobility (Rahman, A, Jerry Situ and Vicki Jimmo, 2005). Encouraging third level education participation has now become a key policy objective for most governments around the world; the participation of young people in higher education has increased significantly in the last twenty years in the majority of developed economies (OECD, 2009). The desire for a highly educated population stems from the belief that education can help economic growth by influencing worker productivity (Mankiw, Romer and Weil, 1992).

Participation means that students are formally enrolled in school. However, this crude description does not take into account the quality of participation a student may have – their ‘engagement’ – with school. A student may be enrolled, but not actively participate at all. This may be because they are not effectively participating because of truancy, or because of their school behavior and the disciplinary procedures associated with that. More commonly, as students in this study have demonstrated, it is because of poor teaching methods and lack of learning resources. Consultations with students discussed in this study revealed frequent reference to this qualitative dimension of participation as being very important to them and a factor strongly related to participation and transfer (Robert, C., and Arlianti, R., 2009). According to the participation statistics, the categories which count towards participation in education and training today are full-time education (FTE), work-based learning (WBL), and ‘other education and training’ (OET). (Mark Corney, 2009). But also, in the context of international goals and commitments, the number of out-of-school children is one of the most frequently cited education statistics. It is therefore crucial that, not only an appropriate definition and methodology are used, but that there is a good understanding of the results, their interpretation and limitations. It is important to clearly recognize that the final goal is not only to get children in school but to ensure schooling results in good learning outcomes. (UNESCO, 2005).

Education brings wide-ranging benefits to both individuals and societies. It is considered so important to individual development that the right to primary education is legally guaranteed in most countries of the world. Moreover, international human rights conventions also recognize the right to education. This right has been established by a succession of UN Conventions, from the Universal Declaration of Human Rights (1948) to the Convention on the Rights of the Child (1989), which acquired the status of international law in 1990. According to Article 28 of the Convention, governments have the responsibility of making primary education compulsory and available free to all. Education is also recognized as crucial to human development. Indeed the Education for All (EFA) movement and the Millennium Development Goals (MDG) have led to greater attention paid to educational participation and completion. (UNESCO, 2005). To understand the characteristics of Education Participation Index, there are many approach has been done to analyze the Education Participation Index, for example the application of a multivariate regression analysis was used to identify the net effects on the likelihood of school attendance for each of the five variables. In the model, the dependent variable is current school attendance, and the independent variables are age, sex, mother's education, household wealth and place of residence, the model was tested with a logistic regression (UNESCO, 2005); the application of multiple regression with dummy variable (Jean Drèze and Geeta, G.K.,1999) to analyze School Participation in Rural India; The application of logit model also has been used (Rahman, A, Jerry Situ and Vicki Jimmo, 2005) to analyze Participation in Postsecondary Education: Evidence from the Survey of Labour and Income Dynamics.

In this paper we will try to analyze data of Education Participation Index in Indonesia from 2003 to 2008, the data was taken from Central Bureau Statistics of Indonesia (BPS). The data was classified based on the level of ages, regions, and years. To see the differences of the characteristics EPI based on the classifications above, the data will be analyzed by nested design approach, and to modeling the data will be analyzed by multiple regression with some variables are dummies variables.

2 ANALYSIS AND MODELING OF THE EPI

To analyze the data, we used analysis of nested design where the level of ages (7-12, 13-15, 16-18, and 19-24 year-old) are nested within the year (2003 to 2008). The regions are divided into six regions, namely: Sumatra(S)(Aceh, North Sumatra, South Sumatra, West Sumatra, Riau, Jambi, Bangka-Belitung Island, Bengkulu and Lampung), Java(J)(Jakarta, West Java, Banten, Mid Java, Yogyakarta, and East Java), Bali (B) (Bali, West Nusa Tenggara, East Nusa Tenggara), Kalimantan(K)(West Kalimantan, Mid Kalimantan, South Kalimantan, East Kalimantan), Sulawesi(SL)(North Sulawesi, Gorontalo, Mid Sulawesi, South Sulawesi, West Sulawesi, East West Sulawesi), Papua(P)(Maluku, North Maluku, Papua, West Papua). Also in this design assume that the Regions, Years, and level of ages are fixed effect. And the model similar to the model given in Mustofa, et al (2008) is as follow:

$$EPI_{ijkm} = \mu + R_i + Y_j + (RY)_{ij} + A_{k(j)} + (RA)_{ik(j)} + \varepsilon_{m(ijk)} \quad (1)$$

Where EPI_{ijkm} is students' participation index at ith-region, jth years, kth level of age and m sub region, μ is general mean, R_i is the effect of the-ith region, Y_j is the effect of years at the jth year, $(RY)_{ij}$ is the interaction effect due to regions and Years, $A_{k(j)}$ is the effect of level of ages nested in years, $(RA)_{ik(j)}$ is the interaction effect of regions and level of ages nested within the years, and $\varepsilon_{m(ijk)}$ is the error. Based on this model, we can analyze the difference of EPI based on the Regions, Years, interaction Regions and Years, difference means of the level of ages and the interaction between level of ages and Regions nested in Years. To modeling the EPI, we use multiple regressions with dummy variables for the Regions and Years. The model is developed based on the model which can be found in Gujarati (1970), Skvarcius and Cromer (1971) and Montgomery and Peck (1992). The model can be developed as follow:

$$\begin{aligned} EPI &= \beta_0 + \beta_1 A_k + \beta_2 A_k^2 + \beta_3 R_1 + \beta_4 R_2 + \beta_5 R_3 + \beta_6 R_4 + \beta_7 R_5 + \beta_8 Y_1 + \beta_9 Y_2 + \beta_{10} Y_3 + \beta_{11} Y_4 + \beta_{12} Y_5 + \\ &\beta_{13} A_k R_1 + \beta_{14} A_k R_2 + \beta_{15} A_k R_3 + \beta_{16} A_k R_4 + \beta_{17} A_k R_5 + \beta_{18} A_k Y_1 + \beta_{19} A_k Y_2 + \beta_{20} A_k Y_3 + \beta_{21} A_k Y_4 + \\ &\beta_{22} A_k Y_5 + \beta_{23} A_k^2 R_1 + \beta_{24} A_k^2 R_2 + \beta_{25} A_k^2 R_3 + \beta_{26} A_k^2 R_4 + \beta_{27} A_k^2 R_5 + \varepsilon \qquad (2). \end{aligned}$$

where EPI_{ijkm} is students' participation index at ith-region, jth years, kth level of age, β_0 is intercept, A_k is the kth age level of students', A^2_k is the square of ages, and

R_1 = 1, if the observation lies in Sumatra(S),
= 0, otherwise.
R_2 = 1, if the observation lies in Java (J),
= 0, otherwise.
R_3 = 1, if the observation lies in Bali (B),
= 0, otherwise.
R_4 = 1, if the observation lies in Kalimantan (K),
= 0, otherwise.
R_5 = 1, if the observation lies in Sulawesi (SL),
= 0, otherwise.

and

Lies in Papua if R1 = $R_2 = R_3 = R_4 = R_5 = 0$.
Y_1 = 1, if the observation lies in the Year 2003,
= 0, otherwise.
Y_2 = 1, if the observation lies in the Year 2004,
= 0, otherwise.
Y_3 = 1, if the observation lies in the Year 2005,
= 0, otherwise.
Y_4 = 1, if the observation lies in the Year 2006,
= 0, otherwise.
Y_5 = 1, if the observation lies in the Year 2007,
= 0, otherwise.

and

Lies in Year 2008, if Y1= $Y_2 = Y_3 = Y_4 = Y_5 = 0$.

3 ANALYSIS, MODELING AND DISCUSSION

From the analysis by nested design, we have that the model is very significant (p-value <0.0001) and the degree of determination is 97.12% (R2 = 0.9712) this mean that 97.12% the variation of EPI can be explained by the model
(Table 1).

Table 1. ANOVA table from nested design

Source	df	Sum of squares	Mean squares	*F* value	Pr>f
Model	143	782296.0126	5470.6015	144.16	<.0001
Error	612	23225.0568	37.9494		
Corrected Total	755	805521.0694			

R-Square = 0.9712

Table 2. ANOVA table from nested design to test the component of the model

Source	df	SS	Mean Squares	*F* value	Pr>f
Region	5	2746.3849	549.2770	14.47	<.0001
Years	5	926.5160	185.3032	4.88	0.0002
Region*years	25	253.6967	10.1479	0.27	0.9999
Age (years)	18	776014.2822	43111.9046	1139.04	<.0001
Age (region) years	90	2355.1329	26.1681	0.69	0.9854

From Table 2. Over all EPI if we compare the EPI by year, it is significant (p-value <0.0001) and Table 3, show that there are positive significant difference between year 2007 (Y5) and Year 2003 (Y1); also year 2008 (Y6) and year 2003(Y1). There are positive increments by 3.15% in 2007 and by 3.36% in 2008, compared to 2003. Other interesting one, even though there are not significant between 2004, 2005 and 2006 with year 2003, but the different are all positive. This reveals that year by year the EPI in Indonesia increase positively (Table 3.). There are significantly different by Regions (p-value<0.0001), and from Table 4. Sumatra and Sulawesi, Java and Sulawesi, Sumatra and Bali, Jawa and Bali, Sumatra and Kalimantan, Java and Kalimantan and Papua and Sulawesi are all significantly different (at level of significant 0.05) while the other Regions are not significantly different.

Table 3. Mean differences of EPI between Years, ($Y_i - Y_j$), i< j, i=2,3,4,5,6 and j=1,2,3,4,5.

Years	Y_2	Y_3	Y_4	Y_5	Y_6
Y_1	1.75	2.03	2.14	3.15**	3.36**
Y_2		0.28	0.39	1.40	1.61
Y_3			0.11	1.12	1.33
Y_4				1.01	1.22
Y_5					0.21

** Significant at level 0.05.

Table 4. Mean differences of EPI between Regions

Region	SL	B	K	S	J
P	-3.55**	-2.22	-1.58	1.01	1.63
SL		1.33	1.99	4.56**	5.18**
B			0.63	3.22**	3.85**
K				2.60**	3.22**
S					0.62

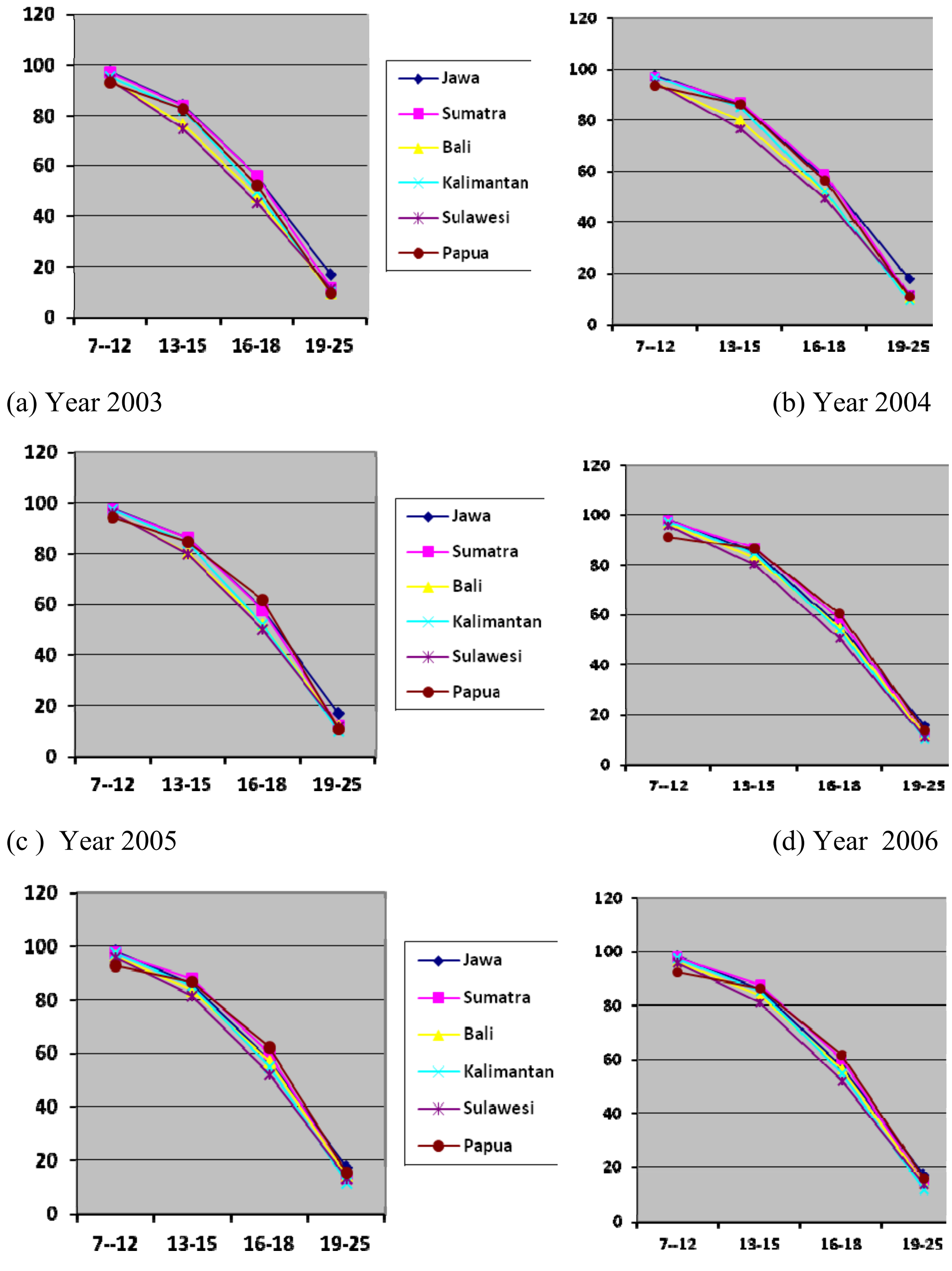
** Significant at level 0.05. J (Java), S (Sumatra), K (Kalimantan), B (Bali), SL (Sulawesi), and P (Papua)

From Table 5. The EPI at the level age 7-12 year-old, the EPI in all region and yeas are above 91%., with the lowest EPI is Papua and Maluku (P). The EPI and its standard deviation are 92.94% and 6.29% in 2003, 93.53% and 5.97% in 2004, 94.16% and 6.79% in 2005, 91.55% and 8.06% in 2006, 92.62% and 6.43% in 2007, and 92.72% and 6.47% in 2008. But in this regions the variation are quite high compared to the other regions, which shows that in this regions the EPI at the level of age 7-12 year-old, distributed not evenly. While in the other regions the variation is small, which shows that the EPI are distributed evenly. The EPI at the level age 13-15 year-old, the EPI are about the in the range 74% to 86% in 2003 and 2004, and about in the range 80% to 87% in 2005 to 2008. But they have the same characteristics, namely they have high standard deviation. This shows that at the level age 13-15 year old in all regions distributed not evenly. The EPI at the level age 16-18 year-old, the EPI in Sumatra in year 2003 and 2004 has highest EPI. The EPI in Papua and Maluku in 2005 to 2008 have the highest EPI. But they have the same characteristics, namely they have high standard deviation. This shows that at the level age 16-18 year old in all regions distributed not evenly, and Java has the highest standard deviation. The EPI at the level age 19-24 year-old, the EPI in Java from 2003 to 2008 has the highest EPI, about 17%, but has the highest standard deviation, which indicates that in Java is distributed not evenly, with the range of between 9% and 47%. Yogyakarta has the highest EPI at the level of age 19 to 24 year-old about 41% to 47% EPI. While in the other regions the EPI are relatively close in the range between 9% and 16% and its standard deviation also relatively small. This shows that at the level age 19-24, out of Java, in the other regions the EPI distributed evenly.

Table 5. EPI and Standard deviation (sd) based on region, years and level of age

	Level of age	7-12		13-15		16-18		19-24	
		EPI	sd	EPI	sd	EPI	sd	EPI	sd
2003	**J**	97.51	0.91	84.37	7.49	55.70	13.37	16.86	12.89
	S	96.94	0.81	83.72	5.83	55.71	9.61	11.89	4.88
	B	94.26	3.28	76.66	8.15	47.59	12.76	9.79	3.59
	K	95.72	2.51	82.12	6.52	49.49	8.67	10.14	1.04
	SL	94.49	2.89	74.48	7.04	45.27	6.90	11.30	2.72
	P	92.94	6.29	82.72	6.61	51.99	3.29	9.71	2.27
2004	**J**	97.67	0.88	86.10	6.45	57.77	12.34	18.29	14.62
	S	96.93	0.69	86.95	4.97	59.01	8.89	11.98	5.41
	B	94.67	1.44	80.10	5.44	51.97	9.86	10.26	0.43
	K	97.04	1.29	85.14	5.47	52.13	8.77	9.45	2.11
	SL	94.48	2.08	76.88	7.02	49.36	8.04	11.92	2.05
	P	93.53	5.97	86.29	7.19	56.73	8.18	11.20	2.25
2005	**J**	97.89	1.04	86.32	7.29	57.94	10.64	17.16	12.24
	S	97.49	0.53	86.29	4.07	57.67	6.21	12.16	3.94
	B	95.91	1.56	80.42	4.21	52.06	8.82	12.30	2.21
	K	97.41	1.31	84.11	7.24	52.05	7.03	10.16	1.38
	SL	95.92	2.03	79.93	7.66	50.18	6.38	11.78	2.91
	P	94.16	6.79	84.57	8.73	61.66	7.41	11.16	1.53
2006	**J**	98.25	0.70	85.36	5.18	55.63	9.29	15.72	12.12
	S	97.64	0.73	87.15	4.53	58.76	8.38	11.96	5.03
	B	96.34	2.16	83.08	5.19	55.11	8.36	11.84	0.99
	K	97.18	0.92	84.46	4.83	53.68	7.25	10.30	1.86
	SL	95.67	1.74	80.37	5.37	50.53	5.69	11.07	2.84
	P	91.55	8.06	86.47	6.15	60.47	7.45	13.82	1.81
2007	**J**	98.37	0.67	85.83	4.98	57.23	8.71	17.31	13.03
	S	97.62	0.84	87.68	4.20	59.79	8.05	13.78	5.30
	B	96.38	2.39	83.64	4.93	56.75	6.91	14.12	0.91
	K	97.58	0.76	85.04	4.85	54.72	6.84	11.78	1.77
	SL	95.80	1.67	81.27	4.67	52.16	5.22	13.54	2.35
	P	92.62	6.43	86.68	5.83	62.08	7.90	15.61	1.88
2008	**J**	98.64	0.63	86.08	4.87	57.29	9.05	17.60	12.95
	S	97.93	0.90	87.84	4.36	59.95	7.94	14.26	4.77
	B	96.47	2.46	83.80	5.38	56.75	6.86	14.17	0.56
	K	97.84	0.67	85.34	4.60	54.84	6.74	11.90	1.72
	SL	96.19	1.59	81.27	4.88	52.31	5.15	13.82	2.27
	P	92.72	6.47	86.54	5.71	61.89	7.69	16.27	1.46

Note: EPI: Education Participation Index (%), sd: Standard deviation



(a) Year 2003 (b) Year 2004

(c) Year 2005 (d) Year 2006

(e) Year 2007 (f) Year 2008

Figure 1. Graph of the EPI by Year, Region and level of age

Table 6. shows the result of statistical test for model (2) and the result shows that the model (2) is very significant with p-value <0.0001 and its degree of determination R^2=0.9700, means that 97% the variation of EPI can be explained by the model, or 97% EPI can be explained by the level of ages, regions and years.

Table 6. ANOVA table for testing the model (2)

Source	df	Sum of Squares	Mean	F Value	Pr > F
Model	27	781384.61	28940.17	872.89	<.0001
Error	728	24136.45	33.15		
Corrected	755	805521.06			

R-Square= 0.9700, Root MSE = 5.75

Table 7, shows the result of parameters estimation and its standard error for the β's parameters in the model (2). From the Table 7. The standard error for parameters related to regions has higher standard error compared to standard error related to Years. This shows that the EPI related to regions distributed not evenly. The EPI as a function of the level of ages for each region and years are given in Table 8. The model can be written as:

$$EPI = \beta_0^* + \beta_1^* A_k + \beta_2^* A_k^2 + \varepsilon \qquad (3)$$

For example, for Java with R_1=1 and =0 otherwise, and year 2004, Y_2 = 1 and =0 otherwise. If we put all these values in to the model estimation of model (2), we get the model for Java in 2004:

$$EPI = 89.40 + 15.50\, A_k - 8.76\, A_k^2 + \varepsilon$$

With the equivalent calculation for other regions and years to get the model (3) for each regions and years and the results are presented in Table 8.

Table 7. Estimation and testing the parameters in model 2.

Parameter	Estimation	Standard Error	t Value	Pr > |t|
Intercept	80.58	3.66	21.98	<.0001
AGE	22.66	3.22	7.05	<.0001
AGE2	-9.76	0.63	-15.54	<.0001
R1	10.43	4.09	2.55	0.0110
R2	17.92	4.40	4.07	<.0001
R3	16.48	5.15	3.20	0.0014
R4	17.04	4.79	3.56	0.0004
R5	21.75	4.47	4.86	<.0001
Y1	-2.07	1.78	-1.16	0.2449
Y2	-0.61	1.78	-0.35	0.7301
Y3	0.06	1.78	0.03	0.9736
Y4	0.29	1.74	0.17	0.8669
Y5	-0.15	1.74	-0.09	0.9299
AGE*R1	-6.74	3.73	-1.80	0.0715
AGE*R2	-15.20	4.01	-3.79	0.0002
AGE*R3	-16.58	4.69	-3.52	0.0005
AGE*R4	-14.61	4.37	-3.34	0.0009
AGE*R5	-23.20	4.08	-5.68	<.0001
AGE*Y1	-0.53	0.65	-0.82	0.4127
AGE*Y2	-0.42	0.65	-0.64	0.5223
AGE*Y3	-0.57	0.65	-0.88	0.3785
AGE*Y4	-0.60	0.63	-0.95	0.3418
AGE*Y5	-0.02	0.63	-0.03	0.9737
AGE2*R1	1.00	0.73	1.36	0.1735
AGE2*R2	2.91	0.79	3.69	0.0002
AGE2*R3	3.06	0.92	3.28	0.0011
AGE2*R4	2.40	0.86	2.79	0.0054
AGE2*R5	4.37	0.80	5.43	<.0001

Table 8. Parameter Estimation model for each Regions and Years

Years	Parameters	Regions					
		J	S	B	K	SL	P
2003	β_0*	88.94	96.43	94.99	95.55	102.26	78.51
	β_1*	14.39	6.93	5.55	7.52	-1.07	22.13
	β_2*	-8.76	-6.82	-6.70	-7.36	-5.39	-9.76
2004	β_0*	89.40	97.89	96.45	97.10	101.72	87.97
	β_1*	15.50	7.04	5.66	7.63	-0.96	22.24
	β_2*	-8.76	-6.82	-6.70	-7.36	-5.39	-9.76
2005	β_0*	91.07	98.56	97.12	97.68	102.39	88.64
	β_1*	15.35	6.89	5.51	7.48	-1.11	22.09
	β_2*	-8.76	-6.82	-6.70	-7.36	-5.39	-9.76
2006	β_0*	91.30	98.79	97.35	97.91	102.62	88.87
	β_1*	15.32	6.86	5.48	7.35	-1.14	22.06
	β_2*	-8.76	-6.82	-6.70	-7.36	-5.39	-9.76
2007	β_0*	90.86	98.35	96.91	97.47	102.18	88.43
	β_1*	15.90	7.44	6.08	8.03	-0.56	22.64
	β_2*	-8.76	- 6.82	-6.70	-7.36	-5.39	-9.76
2008	β_0*	91.01	98.50	97.06	97.62	102.33	88.58
	β_1*	15.92	7.46	6.08	8.05	-0.54	22.66
	β_2*	-8.76	-6.82	-6.70	-7.36	-5.39	-9.76

REFERENCES

Debroucker, P. (2005), *Getting There and Staying There: Low-income Students and Post-secondary Education,* www.cprn.org/, Canadian Policy Research Network (CPRN).

Gujarati,D.(1970). Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Linear regressions: A Generalization. *The American Statistician*, December, 1970.

Jean Drèze and Geeta, G.K. (1999). *School Participation in Rural India*, Institute for Economics and Statistics (University of Oxford).

Mankiw N., D. Romer and D. Weil, (1992). A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics*, Vol. 107.

Mark Corney,(2009). *Raising the Participation age*, CfBT Education Trust, UK.

Mustofa, U., Elfaki, F.A.M., and Jamal, I.D.(2008). *Experimental Design: For Scientists and Engineers*. Kuala Lumpur: IIUM Press.

Montgomery,D.C. and Peck,E.A.(1992). *Introduction to Linear Regression Analysis*, (Second Ed.). New York: John Wiley and Sons, Inc.

OECD, (2009). *Education Today*, The OECD Perspective. OECD Publishing, Paris

Oreopoulos, P. (2006a). The compelling effects of compulsory schooling: evidence from Canada, *Canadian Journal of Economics* , 39, 1, 22-52.

Rahman, A, Jerry Situ and Vicki Jimmo (2005), *Participation in Postsecondary Education: Evidence from the Survey of Labour and Income Dynamics*, Published by authority of the Minister responsible for Statistics Canada © Minister of Industry, 2005

Robert, C., and Arlianti, R. (2009). *Transition to and Participation in Junior Secondary School, Study Case in Indonesia*, The United States Agency for International Development (USAID), Jakarta, Indonesia.

Skvarcius,R., and Cromer, F.(1971). A Note on the Use of Categorical Vectors in Testing for Equality of two Regression Equations. *The American Statistician*, June, 1971.

Shaienks,D., and Tomasz, G., and Justin , B. (2008), Postsecondary Education –Participation and Dropping out: Differences Across University,College and Other Types ofPostsecondary Institutions, *Published by authority of the Minister responsible for Statistics Canada © Minister of Industry, 2008*

UNESCO, (2005). *Children Out of School: Measuring Exclusion from Primary Education*, UNESCO Institute for Statistics, Montreal, 2005