RESEARCH ARTICLE

Human capital formation within families: a study in the North Central Province of Sri Lanka

D P S Chandrakumara
Department of Economics, Faculty of Humanities and Social Sciences, University of Sri Jayewardenepura, Nugegoda

Abstract: The aim of this paper is to uncover the role of family in human capital formation through education, so as to draw important policy implications for poverty reduction. The study used both descriptive statistics and Binary Logistic Analysis to examine the influence of different family factors on the education level of children. The study is based on primary data collected from three villages of the North Central Province of Sri Lanka. The study found that four family factors- namely, family assets, presence of senior siblings, presence of junior siblings and father's age, had been significant in determining the children's education level thereby forming a family role model. The two factors, family assets and presence of senior siblings, had a positive relationship with children's education level, while the other two had an inverse relationship. Further, it was found that gender, living area of the family were not significant factors to be included in the model. Finally, the model developed by the study is useful in the public policy decisions on human capital formation through education, which uses the family as the bottom level basis for reducing poverty.

Key words: Human capital, Education, Family factors, Binary Logistic Model

INTRODUCTION

Background

In a society where children's education from the kindergarten to the collegiate level is mostly controlled by parents and adults, the family plays a key role in human capital formation. Parents and children collectively involve in taking important decisions and allocate limited family resources for education of children and other purposes (Giannelli and Monfardini, 2000; Behrman, Pollack and Taubman, 1995; Blundell, Dearden and Sianesi, 1999). Thus, the demand for education of children is basically dependent upon how far parents and adult family members direct and motivate the children for education using their limited resources.

A family is the first and the most immediate socio-economic environment to which a child is exposed. It is an outstanding primary group, because it is in the family that the child develops his/her knowledge and attitudes. A family can basically exist in two forms, nuclear family and extended family. A nuclear family ordinarily consists of father, mother, one or more children, while an extended family consists of near or distant relatives and even servants (Behrman et al., 1995; Rao, 1990; Horton and Hunt, 1980). The family may get more and more children and become more complicated with the passage of time, so that it may include elder and younger children and male and female children etc. It also undergoes continuous changes in knowledge, experience and attitudes of parents and children. Thus, the family background that a child faces during his/her education process, changes from age to age and also from family to family. Therefore, it seems that the family's role in human capital formation through education is a very complicated process.

Human capital created in individuals due to education is measured by the level of education that they have attained. In that sense, the meaning of human capital formation is the progression of individuals in their path of education from kindergarten to the tertiary levels passing different stages such as primary, junior secondary and senior secondary. Human capital formation through education

chandrakumara62@gmail.com

which occurs at family level is important for an economy in two main ways. First, it builds up the labour force needed for a knowledge-based development and generates social returns (Eicher, 1999; Mankiw, Romer and Weil, 1992; Lucas, 1988). Second, education increases the stock of human capital, which in turn increases labour productivity and wages and reduces poverty (Aakvik, Salvanes and Vaage, 2005; Ranis, 2004; Tilak, 2002a, 2002b; Mingat and Tan, 1996; Weale, 1992; Sen, 1995, 1985; Schultz, 1961). Since labour is by far the most important asset of the poor, provision of education to the poor will tend to reduce poverty through increased labour productivity (Rodriguez, 2002; Guneratne, 1985). Low-education is one of the most serious causes of poverty. In fact, there seems to be a vicious circle of poverty as low education leads to poverty and poverty leads to low education. The poor are not able to afford education, even if it is freely provided by the government, because of the high extra financial and opportunity costs that they face along with the need to sacrifice current consumption (Mingat and Tan, 1996). In a situation where there are no credit facilities available for students for education, the importance of family members becomes still higher (Roussel, 2000). Therefore, this paper attempts to uncover the role of family in human capital formation through education in the North Central Province of Sri Lanka.

Importance of the Study

The study is theoretically and practically important due to two main reasons:

First, the study builds a sound and productive empirical basis on human capital formation at a family level and provides policy implications related to education and poverty reduction in Sri Lanka. It helps to understand how the family units can be used, not only for producing the necessary human capital requirements but also as a strategy for reducing poverty at the family level.

Second, the study adds theoretical knowledge to research on human capital formation within families. It tests the validity of the model that family factors such as parents' occupation, family assets together with sibling structure determine the intra-family human capital formation. In a society where economic activities are largely determined at the family level, this study turns the attention of economists to analyze the family in order to find more avenues to mobilize resources for development and poverty reduction within the family itself.

The Problem

The human capital formation through education is determined by both supply and demand for education in any country. The supply of the education service is mainly determined by government policy, quality and the number of public and private schools. The demand side of education is largely determined at the family level. Even if education is freely provided by the government, in the supplying of supplementary resources at the bottom level, families have to play even more active roles in enrolment of children to schools, choice of the level of education, study-field, motivation and guidance. As such, the educational performance of a child is largely dependent on the practices within the family to which the child belongs. Therefore, the objective of the study is to ascertain the role of family in creating human capital within the family through the demand for children's education in different regional contexts of Sri Lanka.

Aim and Objectives

The aim of the study is to identify the role of family in the determination of intra-family human capital formation through education. Within this overall aim, the specific objectives are as follows:

To identify and measure the effect of parents' wealth in the determination of human capital formation within families.

To identify whether the education of elder siblings generates externality effects on the education of younger siblings.

To identify whether the existence of younger siblings have an effect on the education of elder siblings.

Hypotheses

In order to achieve the objectives of the study, the following hypotheses were drawn from the review of literature and tested with empirical data.

- *Hypothesis-1*

Null hypothesis: The wealth level of family does not have an influence on the education level of children of the family.

Alternative hypothesis: The wealth level of family have an influence on the education level of children of the family.

Rationale: Even in a situation where the government provides free education, the benefits are not equally distributed among children. Only the children with a better economic background have the access to better education services such as high quality schools and more effective private tuition. Therefore, a positive relationship between family wealth and the level of education of children is expected.

- *Hypothesis-2*

Null hypothesis: Existence of elder siblings does not have an influence on the education level of younger siblings.

Alternative hypothesis: Existence of elder siblings has an influence on the education level of younger siblings.

Rationale: Elder siblings can either provide financial assistance or proper advice and guidance to younger siblings based on their own experience, knowledge and contacts. In addition, if the children are needed for work, it is more likely that older children will work and the younger will go to school. Therefore, existence of elder siblings positively influences the education of younger siblings.

Limitations

The study is subjected to a few limitations as follows:

First, the North Central Province (NCP) of Sri Lanka was selected for the study as it was possible to cover the cost of data collection under the SIDA/SAREC Research Cooperation Project for this area.

Second, data collection was carried out in three areas in one district of the North Central Province of Sri Lanka. This district was considered as appropriate because it is the district that consists of both rural and more urban characteristics compared to the other districts in the same province.

Third, the study was confined to the human capital formation through formal education, even though there may be a process of human capital formation through informal education. Formal education is associated with structured learning formats, purpose of obtaining an educational qualification, organized institutions, methods or procedures, etc., while informal education is the knowledge obtained by association, asking questions from co-workers, being shown how to do something and watching others etc. This was arranged in order to avoid the possible complications of the study.

Finally, despite the differences in the stream of study followed by different children – Arts, Commerce and Science – all children who reached Advanced Level were considered as the children who achieved the same position in education.

METHODOLOGY

Specification of the Model

In order to determine the relationship between family characteristics and a child's level of education, a set of explanatory variables were initially identified with the help of the literature review. These variables could be arranged to form the simplified children's education function. Education of child 'i' in j^{th} family was taken as a production function,

$$E_{ij} = f(GNR_{ij}, FA_{ij}, FP_{ij}, MA_{ij}, MP_{ij}, NJS_{ij}, NSS_{ij}, NOM_{ij}, ASI_{ij}, AREA_{ij})$$

where,

E_{ij}	=	Education level of ith child in j^{th} family (Primary level-1, Lower secondary level-2, Higher secondary level-3)-Categorical variable
GNR_{ij}	=	Gender (Male-1, Female-2)-Categorical variable
FA_{ij}	=	Father's Age - Continuous variable
FP_{ij}	=	Father's Profession (Categ-1, Categ-2, Categ-3, Categ-4, Categ-5) - Categorical variable
		Categ-1: No stable job
		Categ-2: Agriculture, animal husbandry and minor self employment
		Categ-3: Minor workers in public and private sectors
		Categ-4: Traders
		Categ-5: Teachers and public and private executives
MA_{ij}	=	Mother's Age - continuous variable
MP_{ij}	=	Mother's Profession (Categ-1, Categ-2, Categ-3, Categ-4, Categ-5) - Categorical variable (Categories are same as in father's profession)
NJS_{ij}	=	Number of Junior Siblings - Discrete variable
NSS_{ij}	=	Number of Senior Siblings - Discrete variable
NOM_{ij}	=	Number of Other Members in j^{th} family of ith child
ASI_{ij}	=	Assets index of the ith child of jth family - Continuous variable
$AREA_{ij}$	=	Area of the ith child of jth family– Categorical Variable

A child's siblings are separated into two groups in order to capture the true interactions among siblings within a family. For child 'i' in j^{th} family, let the number of junior sibings be denoted by NJS_{ij}. Similarly, NSS_{ij} denotes the number of senior siblings.

The economic potential of a family is represented by an asset index due to two main reasons. First, using 'family income' for representing economic potential of a family is difficult, as people generally do not provide accurate information. Second, the current family income may not measure the long-term economic status of a family which is more relevant for the educational achievements. Hence, instead of using 'family income' for representing economic potential of a family, an 'asset index' (ASI_{ij})

was constructed with the help of the statistical technique 'Multiple Correspondent Analysis'.

As the dependent variable is dichotomous, the Binary Logistic Model was used for the study. The structural equation for the dependent variable can be presented in the following form:

$$Y_{ij}^{*} = x_{ij}\beta_{ij} + \varepsilon_{ij}$$

Where Y is the dependent variable, x_i is the set of explicative variables, β is the set of the parameters to be estimated and ε is the error term. The relationship between exploratory and dependent variables can be introduced in the following form:

$$Y_{ij}^{*} = \begin{cases} 2\_if\_Y^* > 1 \\ \\ 1\_if\_Y^* \leq 1 \end{cases}$$

Therefore, for given values of exploratory variables:

$$Pr(y = 1|X) = Pr(y^* > 1|X)$$

Knowing the structural equation for the dependent variable, transforming and re-arranging terms, the above equation can be expressed in the following form:

$$Pr(y = 1|X) = Pr(\varepsilon > -(\alpha + \beta X)|X)$$

Thus, in this model with a dependent binary variable, the probabilities of the child to be successfully enrolled in the Higher Secondary Level are estimated by using the logistic distribution function.

Description and Construction of Variables

The Dependent Variable

The dependent variable (N=131) is a dichotomous variable indicating whether or not a child reached the higher secondary level. The value is 1 if the child did not enter the advanced level. In a conceptual sense, this category includes the children who never attended school and the children who dropped out of school before Advanced Level if they are in the age range of 18-21 years as at the time of the interview. However, practically, there was no child who never attended school. The variable takes the value of 2 if the child has either

registered for an Advanced Level course or completed the course despite its successfulness by the age range of 18-21 years. Table 1 shows how the dichotomous variable, level of schooling, has been distributed in the three villages.

Table 1: Distribution of Children (Age 18-21 yrs) by Level of Schooling

Level of schooling	Kala Medawachchi	Halmillewa	Thammennakulama	Total
Less than A/L	21	19	14	54
A/L	17	11	49	77

The Explanatory Variables

After reviewing the numerous studies completed as of Shultz (1961) on human capital and Becker (1964) on human capital formation within families, at the first step, eleven variables were included in the model. The following is a list of all variables used in the estimated function, as well as the rationale behind their inclusion.

Assets Index (ASI): In order to represent the economic potential of the family, an 'asset index' was constructed. This represents the resources possessed by the family that can influence the child's academic aspirations and success.

Aiming the construction of an assets index, the researcher included three sets of questions related to the economic status of the family. First, families were asked about their ownership of various assets, such as whether they own a radio, a television set, a refrigerator, a bicycle, a motor-cycle, a land vehicle, a motor car. Second, they were asked about characteristics of their housing, namely the number of square feets of their house per head, whether the roof is permanent, whether the wall is permanent, whether the floor is permanent. Third, in order to include the basic human capital assets of the family, the level of education of the father and mother was asked. Altogether, the number of variables constructed through the questions was 12 while the number of level categories was 29 (see Table 2). To ensure comparability across the three areas, only variables that appear in all of the three areas were included in the analysis. As the extent of land owned by the families in the suburb area (Thammennakulama) was much less than that of rural villages, it was not included as a common component of assets. Although the amount of jewellery or gold owned by families was a good component of wealth, it was practically difficult to get information about such assets which is the same as for

income. However, since we have 12 asset type variables, it gives a 12-dimensioned space, which is impossible to imagine by just a simple visualization.

The following equation was used to calculate the asset index.

$$MCA_{Pi} = R_{i1}W_1 + R_{i2}W_2 + \ldots + R_{ij}W_j + \ldots + R_{in}W_n$$

Where MCA_{Pi} is the ith household's composite asset index, R_{ij} is the response of family i to category j, and W_j is the MCA weight for dimension one applied to category j. The statistical software 'MINITAB' was employed to calculate the weights.

Then the main problem was the choosing of weights for each item of assets. For this purpose, there are two efficient and unbiased techniques in statistics namely, 'principle component analysis' and 'multiple correspondent analysis'. In most of the studies, principal components or factor analysis (PCA) is widely used for the construction of asset indices. However, PCA is essentially designed for continuous variables as it assumes a normal distribution of indicator variables. In contrast, multiple correspondence analysis (MCA) makes fewer assumptions about the underlying distributions of indicator variables and is more suited to discrete or categorical variables. Hence, the MCA rather than PCA in constructing the asset index was employed in the analysis. Using the weights obtained through MCA, the assets index was calculated separately for each family so that it was possible to even rank them on the basis of the value of the index. However, as positive values add to the assets index while negative values decrease the assets score, in some families there was a problem of having negative index values. To obtain positive asset values, a value equal to the greatest negative value was added to each of the asset index values, so that the lowest value becomes zero. A further small magnitude was added to make the lowest value non-zero. The transformation entailed adding 0.1785 to the asset index.

Child's gender: The child's gender was considered here as a child characteristic that separates the child by gender difference. Gender is a dichotomous variable that takes the value of 1 for females and 2 for males. Totally, 63 boys and 68 girls were qualified to be selected from the three villages.

Age of parents: Age of parents which is a continuous variable, consists of two variables namely, father's age and mother's age. The age of parents when the informant child is at the age range 18-21 was considered for the

study. This can be considered as a proxy indicator of the modernity of parents, which may be an important factor affecting the education of children.

Table 2: Components of the Assets Index and the Weights Obtained from Multiple Correspondent Analysis

Asset No.	Component/ Variable	No. of Levels	Level Particulars	Weight (Coordinates)
1	Size of house	3	Hsize 1 (< 500 sq.ft)	-1.850
			Hsize 2 (501- 1000 sq.ft)	-0.144
			Hsize 3 (>1000 sq.ft)	0.566
2	Roof condition	2	Rcond 1 (Temporary)	-1.796
			Rcond 2 (Permanent)	0.343
3	Floor condition	2	Fcond 1 (Temporary)	-1.582
			Fcond 2 (Permanent)	0.450
4	Wall condition	2	Wcond 1 (Temporary)	-1.788
			Wcond 2 (Permanent)	0.341
5	Toilet condition	3	Tcond 1 (No permanent toilet)	-2.262
			Tcond 2 (Permanent-Pit)	0.101
			Tcond 3 (Permanent-Flush)	1.074
6	Motor vehicles	2	MV 1 (Does not own)	-0.136
			MV 2 (Owns a motor vehicle)	0.855
7	Land vehicles	2	LV 1 =(Does not own)	-0.017
			LV 2 (Owns a land vehicle)	0.129
8	Motor-cycles	2	MB 1 =(Does not own)	-0.562
			MB 2 (Owns a motor-cycle)	0.588
9	Color televisions	3	TV 1 (Does not own a TV)	-1.009
			TV 2 (Owns a black & white TV)	0.324
			TV 3 = (Owns a colour TV)	1.302
10	Refrigerators	2	Frig 1 (Does not own)	-0.621
			Frig 2 (Owns a refrigerator)	0.713
11	Father's education	3	FE 1 (< Ordinary Level)	-0.508
			FE 2 (Advanced Level)	0.452
			FE 3 (> Advanced Level)	1.127
12	Mother's education	3	ME 1 (< Ordinary Level)	-0.582
			ME 2 (Advanced Level)	0.502
			ME 3 (> Advanced Level)	1.136

Number of senior siblings: This variable includes the number of all male and female children who are older than the informant child. This is a discrete variable that represents the number of matured members that the informant child has in the family.

Number of junior siblings: The number of all male and female children who are younger than the informant child was taken into account. This is a discrete variable which was included with the intention of identifying whether their influence is different from that of elder siblings.

Number of other members: This variable was included in the model with the purpose of separating the nuclear family from the extended family. The members who are permanently living in the same household, but not the members of the nuclear family, for instance, father-in-law, mother-in-law, any other relative or non-relative residents, were included in this variable. This is a discrete variable.

Residence of the family (Area): In order to specify the residence of the family, dummies for three areas

were included. The nominal values of 1 for Kala-Medawachchiya (traditional rural area), 2 for Halmillewa (rural settlement area) and 3 for Thammannakulama (suburb area) were assigned in order to distinguish the effect of residence.

Parental occupation: Two dummy variables were created to determine the parental professional status. First of all, parental occupation was split into father and mother. Next, based on the type of their main economic activity, each category was split into five nominal variables, viz., (i) No permanent job, (ii) Agriculture and animal husbandry (own) (iii) Mason/carpentry, labourer, minor workers, drivers, armed forces including the police, (iv) Business (v) teaching, government and private sector executives. The aim of dividing the parents' occupation into groups was to approximate the nature of economic activities of the family.

Data Collection

The estimation requires data that provide information on the characteristics of children, parents and others. The primary data was collected through a household survey. A sample of 131 children from 131 families was drawn from three villages in the North Central Province (NCP) of Sri Lanka.

The sampling and data collection procedure followed several steps. **First**, the North Central Province (NCP) was selected for the study because of three reasons: (i) as part of a bigger project that functioned at NCP; by choosing this province, it was possible to get advantages, both scientific as well as economic; (ii) ability to manage data collection process personally, as it was required due to the sensitive nature of the subject; (iii) the possibility for generalization; being a province which has both rural and urban characteristics within the same district, it is possible to generalize the results for a large part of the population of the country. **Second**, Anuradhapura was selected as it is a district that adequately represents different aspects that needed to be covered by the study such as urban-rural divide, high-low family income, different professions, schools of different grades. However, the other district, Polonnaruwa in the NCP did not have any characteristics that varied from the Anuradhapura district. **Third**, two Divisional Secretariat Divisions (DSDs) of the Anuradhapura district, Galnewa and Nuwaragam Palatha (East) were selected, as Galnewa represents rural characteristics while Nuwaragam Palatha (East) represents urban and suburb characteristics. **Fourth**, Thammennakulama of Nuwaragam Palatha (East) was selected because of its non-extremist suburb characteristics: not highly commercialized or sophisticated. However, from the Galnewa DS division

that represents rural families, two G.N. divisions - Karuwalagaswewa and Kala-Madawachchiya - were selected to represent 18 rural settlement G.N. divisions and 12 rural traditional G.N. divisions respectively. **Fifth**, one village from each G.N. division (Halmillewa and Kala-Madawachchiya) was selected depending on the appropriateness of the sampling. **Finally**, all the children, in the age range of 18-21, of selected GN Divisions were identified. The age range was determined at that level because it was necessary to identify children who reached the Advanced Level class and children who dropped out before reaching this level. Accordingly, by that age range, the children either must have attained Advanced Level, if they continued schooling, or dropped out so that the full length of their education up to Advanced Level can be practically identified. As the number of children who belong to the required age range is not so high in all three villages, all the children who were within that range were taken as informants.

Table 3: Size of Sample by Area

Area	Area characteristics	Number of children aged 18-21	Number of children interviewed
1. Thammennakulama	Suburb	57	41
2. Halmillewa	Mahaweli settlement	37	30
3. Kala-Medawachchiya	Traditional village	78	60
Total		172	131

Data collection from the selected children and their parents was completed by three visits to houses in each village. **At the first visit**, families with children in the required age range, the number of children available in each house and other basic information were collected. Moreover, appointments to meet them for the second round were also arranged through the first visit. **The second visit** was arranged only for the families where there were children in the required age range. All the complicated information needed for the quantitative analysis regarding the informant child and his/her family was collected during second visit by using a semi-structured questionnaire.

The semi-structured questionnaire used in the second visit was divided into two parts: PART–A and PART–B. PART-A was to cover family information while PART–B collected information with regard to the child in question. A pilot survey was carried out in each area to test the questionnaire prior to final data collection and then adjusted to test the content of the data collection instrument. Finally, the sample consisted of 131 children who were aged between 18 to 21 and do not have serious disabilities, live with their biological parents, have full information on individual and family characteristics up to Advanced Level education and represent the three main types of areas in the North Central Province of Sri Lanka.

Figure 1: Sampling Procedure
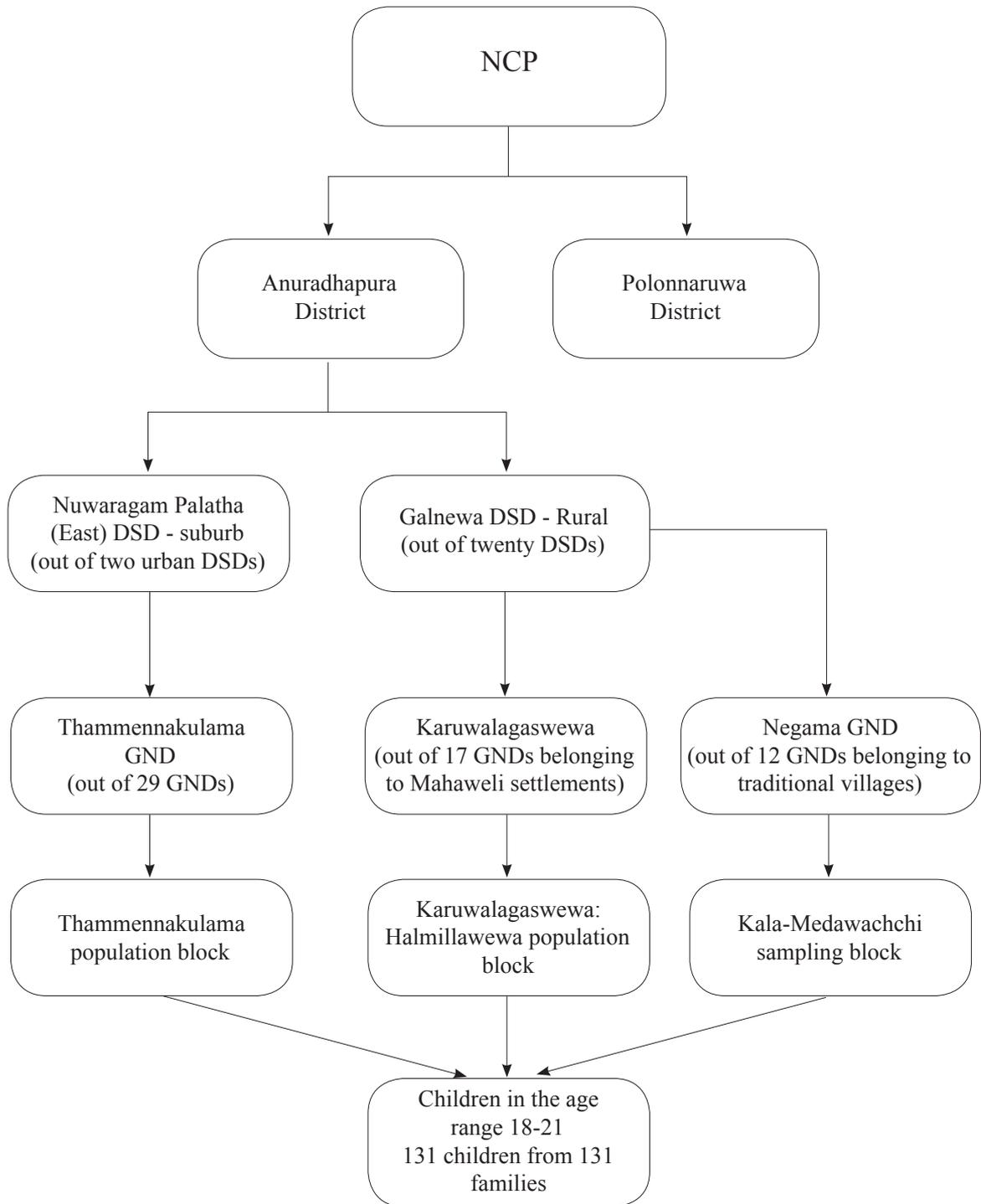
ANALYSIS AND RESULTS

Identification of Relevant Variables (through correlation tests)

In order to uncover the relevant factors for the model, the researcher examined the correlation that the level of education of children have with each explanatory variable which was initially considered for the model. As the dependent variable and most of the explanatory variables are categorical, Spearman Correlations were tested. At this step, eleven exploratory variables were included for the correlation test. This revealed that out of all the variables, the two variables, 'gender of the child (GNR)' and 'number of family members (NFM)' had no correlation with the education level of children to at least 5 per cent level of significance. Hence, the remaining nine variables FA_{ij}, MA_{ij}, FP_{ij}, MP_{ij}, NJS_{ij}, NSS_{ij}, NOM_{ij}, ASI_{ij}, $AREA_{ij}$ (ROF_{ij}) are correlated with the dependent variable at the range of 1– 5 per cent level of significance ($p < .01$ - .05).

Minimizing Multicolinearity (through correlation tests)

When the data is the result of an uncontrolled experiment or primary data collection, it may cause many of the different variables to move together in systematic ways. When this is the case, the variables are said to be collinear or multicollinear when several variables are involved in the econometric testing. This will impose a problem when evaluating the results, since it may not be possible to capture the economic relationship or the parameters of interest. It may also cause lack of significance of individual independent variables while the overall model may be strongly significant. Moreover, it results in wrong signs and magnitudes of regression coefficient estimates and consequently in incorrect conclusions about relationships between independent variables. These problems are highly related to this study since it involves a number of different variables. To eliminate multicollinearity, one way is to represent the highly correlated variables by one of the variables. Another way is to transform the variables to eliminate multicollinearity.

The simple and clear-cut way to detect collinear relationships is to test for the correlation coefficients between pairs of variables. In order to filter out multicollinearity in this study, firstly the correlation coefficients between each independent variable was calculated. In the cases where there are variables with a high degree of collinearity, only one of those variables

was selected so that it represented the other variable/s. However, the variables essential for the model were included in the model despite the collinearity. Although there is collinearity between NJS_{ij} and NSS_{ij}, neither of the variables were removed from the model because it was necessary to include both junior and senior siblings for a model of family. Even if another two variables, ASI_{ij} and NJS_{ij} show multi-collinearity to a certain extent, these two were also brought forward to be included in the model because it was necessary for the model to be meaningful. The two variables, FP_{ij} and MP_{ij} were removed from the analysis at this stage because those variables can be properly represented by ASI_{ij} without a problem to the family model. Since there was multi-collinearity between father's age and mother's age, the latter (mother's age) was removed from the analysis and represented by the father's age.

Binary Logistic Analysis and Results

For the logistic regression, the level of education was recorded into either A/L or below that, since binary response was required. The case processing summary (Table 4) displays the total number of cases or observations included in the analysis as 131 and the number of missing cases as zero. The dependent variable encoding in Table 5 shows the original and internal values of the dichotomous variable. Although the original value that has been given for the dependent variable is 1 if the child in question is below A/L, and 2 if he/she reached A/L, the internal values of the model are respectively 0 and 1.

Table 4: Case Processing Summary

Unweighted Cases (a)	N	Percent
Included in analysis-selected cases	131	100.0
Missing cases	0	0
Total	131	100.0
Unselected cases	0	0
Total	131	100.0

a If weight is in effect, see classification table for the total number of cases.

Table 5: Dependent Variable Encoding

Original Value	Internal Value
1 (Below A/L)	0
2 (A/L)	1

The model building process included six variables at the first step namely, father's age (FA), total number of other members (TOM), assets index (ASI), area

(AREA), number of senior siblings (NSS) and number of junior siblings (NJS) of the child in question. Among these, 'AREA' was a categorical variable that has three categorical codes such as 1, 2 and 3 for the three sampling areas, Kala-Medawachchiya, Halmillewa and Thammennakulama respectively (Table 6).

Table 6: Categorical Variables Codings

		Frequency	Parameter coding	
			(1)	(2)
AREA ij	1	30	1.000	.000
	2	38	.000	1.000
	3	63	.000	.000

Predictive variables were added in a forward stepwise method, known as the Wald method. The significance of the models in each step is evaluated using Chi-square statistics (Table 7). A variable has to be significant at 0.10 level before it can enter into the model, while it has to be significant at 0.15 for it to remain in the model. The Table shows that the models in each step up to step 4 are significant (Sig < 0.01).

Table 7: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	52.651	1	.000
	Block	52.651	1	.000
	Model	52.651	1	.000
Step 2	Step	29.550	1	.000
	Block	82.201	2	.000
	Model	82.201	2	.000
Step 3	Step	8.720	1	.003
	Block	90.920	3	.000
	Model	90.920	3	.000
Step 4	Step	4.283	1	.038
	Block	95.204	4	.000
	Model	95.204	4	.000

The R squared in ordinary multiple linear regression provides an indication of how much variation in the observed values of the predicted variable is explained by the model. However, the Cox & Snell R squared is a form of R squared which arises in logistic regression, and the Nagelkerke R squared is a transformation of the Cox & Snell R squared so that it lies in the interval between zero and one, as is the case in the ordinary multiple regression R squared.

Table 8: Model Summary

Step	−2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	124.895[a]	.331	.466
2	95.345[b]	.466	.628
3	86.625[b]	.500	.674
4	82.342[b]	.517	.696

a. Estimation terminated at literation number 5 because parameter estimates changed by less than .001.

b. Estimation terminated at literation number 5 because parameter estimates changed by less than .001.

The Nagelkerke R square (Table 8) shows that about 45 per cent of the variation in the outcome variable (probability to reach or not to reach to A/L) is explained by the model in step 1, while 70 per cent of this variation is explained by this logistic model in step 4.

The Hosmer-Lemeshow test displays (Table 9) how closely the observed and predicted probabilities match. It provides a formal test to find whether the predicted probabilities match the observed probabilities. This test should be greater than 5 per cent to indicate a good fit based on the difference between observed and predicted frequencies. A larger p-value indicates a better match or insufficiency of evidence to claim that the model does not fit the data adequately. If the p-value is less than the accepted α-level, the test would reject the null hypothesis of an adequate fit. Accordingly, the model is best at the third step when the 'number of junior siblings' are not included. However, as the model should explain the role of different members of the family, the researcher decided to include the 'number of junior siblings' also into the model. Therefore, the model at step 4 was taken as appropriate in explaining the human capital formation within the family.

Table 9: Hosmer and Lemeshow test

Step	Chi-square	df	Sig.
1	24.132	8	.002
2	12.823	8	.118
3	3.471	8	.901
4	9.052	8	.338

In addition, Table 8 gives a classification for the models fitted, and illustrates the predictive accuracy of the logistic regression models from step 1 to step 4. It shows that the model in step 4 was able to predict '*not reaching A/L*' 83.3 per cent, and '*reaching A/L*' 90.9 per cent correctly, while the overall success rate in predicting the children reaching 'A/L' was 87.8 per cent. This result shows that the model, with prediction variables FA, NSS, NJS and ASI, does a good job at explaining the variability amongst individuals with regard to reaching 'A/L'.

Table 10: Classification table

Observed			Predicted		
			Eij		Precentage Correct
			1	2	
Step 1	E$_{ij}$	1	36	18	66.7
		2	9	68	88.3
	Overall Percentage				79.4
Step 2	E$_{ij}$	1	42	12	77.8
		2	8	69	89.6
	Overall Percentage				84.7
Step 3	E$_{ij}$	1	42	12	77.8
		2	7	70	90.9
	Overall Percentage				85.5
Step 4	E$_{ij}$	1	45	9	83.3
		2	7	70	90.9
	Overall Percentage				87.8

a. The cut value is .500

Furthermore, each estimated coefficient of the logit model was tested using the Wald statistics for being significantly different from zero. The significance should be less than 5 per cent. The Wald statistics is calculated by dividing the coefficient (B in Table 11) by its corresponding standard error, and squaring the result. It provides an indication of whether or not the coefficient is significant for the model. The Wald statistic and its significance given in Table 11 indicate that all fitted coefficients are significant (sig < 0.05).

Exp (B) represents an odds ratio. In general, it measures the extent to which the odds in favour of a positive response are raised when the level of the associated explanatory variable is raised from the reference level to the level specified in the table of results. In this study, it measures the extent to which the odds in favour of reaching A/L are raised when the levels of each explanatory variable is raised from the reference level (low) to the highest level (high). Thus, for this study, firstly, it would seem that a high level of NSS is very significant in increasing younger children's retention in education. This conclusion is supported by the 95 per cent confidence for Exp (B) of 3.43 with the limits of 2.27 (lower) and 8.64 (upper). Secondly, the important factor is the ASI that shows an odds ratio of 1.38 within the limits 1.21 (lower) and 1.59 (upper). The two variables, FA and NJS take the third and fourth places respectively when it is determined in terms of the odds ratio.

The Wald statistic has a Chi-squared distribution on degrees of freedom indicated in the column headed 'df'. The column headed 'Sig.' is a p-value associated with the Wald statistic. Table 11 shows that all the variables, NSS, ASI, FA and NJS are significant at 5 per cent level, whereas NSS and ASI are significant at 1 per cent level. This means that NSS and ASI are highly significant.

The logistic regression table, Table 11 (variables in the equation) shows the estimated coefficients, standard error, Wald statistic, z-values, p-values, odds ratio and 95 per cent confidence interval for the odds ratio. From the output, it can be seen that the estimated coefficients for FA, ASI, NJS and NSS have p-values less than 0.05, indicating that there is sufficient evidence that the coefficients are not zero using an α-level of 0.05. The estimated coefficient of -.104 for father's age represents the change in the log of p (< A/L) / p (A/L) with a 1 unit (one year) increase in father's age, with the other factors held constant. The estimated coefficient of -.609 for number of junior siblings represents the change in the log p (< A/L / p (A/L) with a 1 unit increase in junior siblings, with the other factors held constant. It shows that the effect of both of these factors on the level of

education is negative. The estimated coefficient of 1.489 for the number of senior siblings represents the change in the log of p (< A/L / p (A/L) with a 1 unit increase in senior siblings, with the other factors held constant. The estimated coefficient of .327 for the assets index represents the change in the log of p (< A/L / p (A/L) with a 1 unit increase in the assets index, with the other factors held constant. Both of these factors have a positive effect on the education of children.

Although there is evidence that the estimated coefficients for ASI is not zero, the odds ratio is 1.39 indicating that a one unit increase in ASI increases a child's education by 1.39 times. The highest positive contributor to a child's education in a family is the NSS. The odds ratio shows that the increase of one additional senior sibling of a family increases a child's education by 4.43 times. For FA and NJS, the negative coefficients and the odds ratios indicate that one unit (one year) increase in father's age and one unit increase in junior siblings, decrease the child's education by 0.90 and 0.54 respectively.

Table 11: Dependent Variable Encoding

		B	S.E.	Wald	df	Sig.	Exp (B)	95.0% C.I for EXP (B)	
								Lower	Upper
Step 1[a]	ASlij	.307	.056	30.101	1	.000	1.360	1.218	1.517
	Constant	-3.853	.809	22.705	1	.000	.021		
Step 2[b]	NSSij	1.392	.300	21.490	1	.000	.4.024	2.233	7.248
	ASlij	.347	.066	27.325	1	.000	1.415	1.242	1.612
	Constant	-5.693	1.090	27.289	1	.000	.003		
Step 3[c]	FAij	-.118	.045	6.763	1	.009	.889	.814	.971
	NSSij	1.652	.343	23.211	1	.000	5.216	2.664	10.213
	ASlij	.339	.069	24.292	1	.000	1.403	1.226	1.606
	Constant	.150	2.354	.004	1	.949	1.162		
Step 4[d]	FAij	-.104	.048	4.671	1	.031	.901	.820	.990
	NJSij	-.609	.304	4.028	1	.045	.544	.300	.986
	NSSij	1.489	.341	19.104	1	.000	4.432	2.273	8.642
	ASIij	.327	.070	21.946	1	.000	1.387	1.210	1.590
	Constant	.433	2.496	.030	1	.862	1.542		

a. Variable (s) entered on step 1 : ASlij
b. Variable (s) entered on step 2 : NSSij
c. Variable (s) entered on step 3 : FAij
d. Variable (s) entered on step 4 : NJSij

In a logistic regression it is not the actual education level that is predicted, but rather a function, known as the logit function, which is predicted. If this function may be represented by E (p) then as indicated in Table 11, the equation for step 4 in the logistic regression is:

$$E(p) = .433 - .104FA - .609NJS + 1.489NSS + .327ASI$$

In order to use the model to predict whether an individual is reaching or not reaching A/L based on their scores on NSS, NJS, ASI and FA, the equation first required transformation. The logit function E(p) is a function of p, where p indicates reaching or not reaching A/L. The function is given by log (p/1-p). Transforming the equation, for model 4 the regression equation is as follows:

$$p = \frac{e^{.433-.104FA-.609NJS+1.489NSS+.327\,ASI}}{1 + e^{.433-.104FA-.609NJS+1.489NSS+.327ASI}}$$

In the equation, p represents the probability of the binary response being 1, i.e. of reaching A/L. Based on Table 11, if p is greater than 0.5, it indicates reaching A/L, while a 'p' of less than 0.5 indicates not reaching A/L. The results in this section provide support for the two hypotheses indicating that high and low scores on reaching A/L or the level of education can be predicted with statistical significance by the scores on the NSS and ASI and NJS, FA.

The -2 Log Likelihood (-2LL) is a measure of badness-of-fit, illustrating error remaining in the model after accounting for all independent variables. The -2LL of 1567.256 indicates that there is no significant error remaining in the model. The Hosmer Lemeshow Goodness-of-Fit test is the most useful measure of test which assesses the overall model by testing the null hypothesis that all logistic regression coefficients besides the constants are zero. The model's Hosmer Lemeshow Goodness of Fit chi-square is 9.052 with a significance of .338, indicating a good fit.

The results of the quantitative analysis further confirm most of the findings of other studies which were reviewed in the literature on this subject. In addition, the results show that there is a special pattern that creates human capital through education within a family. The family assets, which mostly consists of parents' assets and senior children positively affect the education of children while junior children make a negative impact on the education of senior children within a family. Further, the analysis finds that the parents' age has a negative relationship on their children's education level. This implies that the children of older parents would not continue education unless they are positively affected by senior siblings or family assets.

CONCLUSIONS

Testing the two hypotheses of the study was helpful in achieving the objectives of the study and to address the research question effectively.

The null hypothesis in hypothesis 1, *'The wealth level of family does not have an influence on the education level of children of the family'*, was rejected. While accepting the alternative hypothesis, it suggests that family wealth works as an effective factor that uplifts the intra-family human capital. The empirical model suggests that a one unit increase in ASI increases a child's education by 1.39 times. The family assets level was also closely related to the parents' education level and the occupation.

The null hypothesis in Hypothesis II, *'Existence of elder siblings does not have an influence on the education level of younger siblings'* was also rejected. While accepting the alternative hypothesis, it suggests that *'Existence of elder siblings causes the increase of the education level of younger siblings'*. Furthermore, the results suggest that the existence of senior siblings is the strongest factor that determines the human capital formation within a family. The logistic analysis shows that when it is analyzed by the odds ratio, one unit increase in senior siblings of a family increases a child's education by 4.43 times.

Furthermore, while testing Hypothesis II, it was possible to identify that positive effects are created from senior to junior siblings, while negative effects are created from junior to senior siblings. According to the Logistic Analysis, the odds ratio indicates that a one unit increase in junior siblings decreases the education of senior siblings by 0.54.

In addition, to the above conclusions drawn for the variables which had a direct relationship with the hypotheses, the following conclusions are also important in understanding the role of family in human capital formation:

a) Father's age, which represented parents' age, shows a negative association with the education level of children. A one year increase in father's age decreases the child's education by 0.90.

b) It also reveals that children's education level has no significant association with the rural urban locality.

Finally, it seems that the role of family in human capital formation through education is highly effective and crucial. As an economic unit in the society, the role it plays cannot be replaced by any other way as it has the basic responsibility in controlling the behaviour of children including their education. Hence, the model found by the study is useful in the public policy decisions on human capital formation through education. The study shows how family factors can be used in forming human capital within families if the government uses this strategy for reducing poverty.

References

1. Aakvik, A., Salvanes, K. G. & Vaage, K., 2005. *Education Attainment and Family Background* Discussion Paper. [online] Available at: <http://www.blackwellsynergy.com/doi/abs/10.1111/ j.14 68> [Accessed 12/02/2006].

2. Becker, G.S., 1964. *Human Capital: A Theoretical and Empirical Analysis*. New York: Colombia Press.

3. Behrman, J. R., Pollack R. & Taubman, P., 1995. *From Parents to Child: Intrahousehold Allocations and Intergenerational Relations in the United States*. Chicago: University of Chicago Press.

4. Blundell, R., Dearden, L. & Sianesi, B., 1999. Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy. *Fiscal Studies*, 20(1), pp. 1-23.

5. Eicher, T. & Garcia-Penalosa, C., 1999. *Inequality and Growth*: *The Dual Role of Human Capital in Development*. [online] Available at: <http://www.ideas.repec.org/p/fth/aixmeq/99a47.html> [Accessed 24/09/2002].

6. Giannelli, G. C. & Monfardini, C., 2000. *Joint Decisions on Household Membership and human capital Accumulation of Youths: The role of expected earnings and labor market rationing*. [online] Available at : <http://www.springerlink.com/index/P840JXSQ5A4UTQAM.pdf> [Accessed 24/09/2002].

7. Gunaratne, L., 1985. *Human Capital and Distribution of Personal Income*. Colombo: Central Bank of Sri Lanka.

8. Horton, P.B. & Hunt, C.L., 1980. *Sociology*. U.S.A: McGraw-Hill Inc.

9. Lillard, L.A. & Reville, R.T., 1999. *Life Cycle Human Capital Investment: New Evidence on an Old Issue.* [online] Available at: <http://www.tau.ac.il/~weiss/laborma/HESEDU1001www.pdf> [Accessed 06/04/2002].

10. Mankiw, N.G., Romer, D & Weil, D. N., 1992. A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics*, 107(2), pp. 407-37.

11. Mingat, A. & Tan, J. P., 1996. *The Full Social Returns to Education: Estimates based on Countries Economic Growth Performance*. [online] Available at: <http://www.adb.org/Documents/Books/Lao_Education_Sector_Dev/biblioref.pdf> [Accessed 12/01/2002].

12. Ranis, G., 2004. *Human Development and Economic Growth*, Center Discussion Paper No. 887, Economic Growth Center, Yale University, New Haven. [online] Available at: <http://www.econ.yale.edu/egcenter/> [Accessed 12/02/2006].

13. Rao, C.N.S., 1990. *Sociology: Principles of Sociology with an Introduction to Sociological Thought*. New Delhi: S. Chand and Company Ltd.

14. Rodriguez, J.G., 2002. *The Determinants of Poverty in Mexico*. [online] Available at:<]http://www.blackwell-synergy.com/doi/abs> [Accessed 26/12/2003].

15. Schultz, T. W., 1961. Investment in Human Capital. *The American Economic Review*, 1(2), pp. 1-17.

16. Sen, A. K., 1985. *Commodities and Capabilities*. North Holland: Oxford University Press.

17. Sen, A. K., 1995. Gender, Inequality and Theories of Justice. In: Nussbaum and Glover eds. *Women, Culture and development: a Study of Human Capabilities*. Oxford: Clarendon Press, pp. 259-273.

18. Tilak, J. B. G., 2002. *Building Human Capital in East Asia: What Others Can Learn*. Washington: The World Bank.

19. Weale, M., 1992. *Education, Externalities, Fertility and Economic Growth.* Washington: World Bank. Working papers:WPS (1039).