UNIVERSITY OF NAIROBI

SCHOOL OF MATHEMATICS

Modeling Factors Affecting Academic Transition In Kenya

By

Mercy W. Maina

A Research Project Submitted In Partial Fulfillment of the Requirement of the Award of the Degree of Master of Science in Social Statistics

June 2014

DECLARATION

This project as presented in this report is my original work and it has never been submitted for a degree in any other university.

Mercy W Maina **Reg. No.: I56/79408/2012**

Signature ………………………… Date: ……/…………/2014………………….

This project has been submitted for examination with our approval as university of Nairobi supervisors

Dr. John .M. Ndiritu

Signature…………………………………….Date:……/……/2014……………………

Dr. George. Muhua

Signature…………………………………….Date:……/……/2014……………………

DEDICATION

I dedicate this research project first to the Lord Almighty for giving me the grace to come this far. To my dear husband George Ndirangu and my lovely children Claire and Leon for your immense support during the course of my study. I highly appreciate the financial and moral support you accorded me. I wish you all the best in all your endeavors and especially your career aspirations

ACKNOWLEDGEMENT

I am grateful to the University of Nairobi and especially my supervisors Dr. John Ndiritu and Dr. George Muhua for their guidance and knowledgeable contribution in making this project a success, special thanks to Dr. N. Owuor for his moral and technical support. Special gratitude goes to my fellow colleagues Mary M, Sheila O., D. Mwai, F. Mbugua, R. Wambugu, A. Karigi and S. Kinyua for being so supportive and resourceful throughout this course. Mrs. J. Muriu, thank you for providing linguistic support.

ABSTRACT

Academic transition has been known to be affected by many factors. This project sought to investigate the most effective factors to academic transition. First, the factors were extracted from literature. They were then grouped into demographic factors, social economic factors, social cultural factors, student factors, curriculum and school factors, environmental factors, and social physical factors. Each factor had a number of variables identified from literature. To understand the factors, the study is based on other research findings on the factors affecting transition. The factors sampled out from literature and the data on identified factors were extracted from Kenya Demographic Health Survey of 2008. Using Principal component analysis, the above factors yielded 11 principal components as the most effective factors to academic transition. These were; social economic status, family position, home environment, family composition, regional influence, parents occupation, parents education, house wife status, mother's type of earning, preventive health measures, and ethnicity. Later, the identified principal components were used as predictors to a multiple regression equation with the highest academic level as the response variable. Among the 11, home environment, mother's type of earning and preventive health measures was not significant in estimating academic level. The most effective factors affecting academic transition were regional influence, social economic status, parents' education, house wife status, family composition, ethnicity, family position, and parents occupation, all ranked in levels of effect to transition.

TABLE OF CONTENTS

DECLARATION ii
DEDICATION iii
ACKNOWLEDGEMENT iv
ABSTRACT v
ACRONYMS ix
LIST OF SYMBOLS ix
CHAPTER ONE: INTRODUCTION 1
1.1 Background 1
1.2 Problem Statement 2
1.3 Research Objectives 2
1.3.1 Primary Objectives 2
1.3.2 Secondary Objectives 2
1.3.3 Research Questions 2
1.4 Justification and Significance of the Study 3
1.5 Outline 3
CHAPTER TWO: LITERATURE REVIEW 4
2.0 Introduction 4
2.1 First Grade Transition 4
2.2 Primary School Transition to Secondary School 4
2.3 Parental Impact to Academic Transition 5
2.4 Social, Personal, and Academic Factors Affecting Transition 6
2.5 Physical Development Effect on Academic Performance 7
2.6 Gender Impact on Schooling Transition 8
CHAPTER THREE: METHODOLOGY 10
3.1. Research Design 10
3.2. Study population 10
3.3. Model Formulation 10
3.3.1 Multivariate analysis 11
3.3.2 Demonstration of Variable Correlation 13
3.3.3 How to Interpret a Correlation Matrix 14

3.4 Principal Component Analysis 15
3.4.1 Theory of Principal Component Analysis (PCA) 15
3.4.2 Algebraic Basis of Principal Component Analysis 16
3.4.3 Difference between PCA and Exploratory Factor Analysis 17
3.4.4 Calculating Principal Components 17
3.4.5 Principal Component Analysis Procedure 20
3.5 Multiple Linear Regression Models 25
4.1 Variables Description 26
4.2 Principal Component Extraction 27
4.3 Regression of component scores with education level 33
CHAPTER FIVE: CONCLUSION AND RECOMMENDATION 37
REFERENCES 38
APPENDICES 40
Appendix 1 40
Appendix 2 41
Appendix 3 42
Appendix 4 43

LIST OF TABLES

Table 3.1: Demonstration Of Correlation Matrix 14

Table 4.1: The component Score…………………………….……………………………30

Table 4.2: Rotated Scores...31

Table 4.3 : Table of Extracted Principal Components 32

Table 4.4: Anova Table 33

Table 4.5: Regression Model Summary 33

Table 4.6: Coefficient Estimates...35

Table 4.7: Multiple Regression Mode Interpretation……………..………………………..37

ACRONYMS

KDHS	Kenya Demographic Health Survey
PCA	Principal Component Analysis

LIST OF SYMBOLS

β	Coefficient estimate
σ	Variance
μ	Mean
R	Correlation matrix
S	Covariance matrix
λ	Eigenvalue

CHAPTER ONE: INTRODUCTION

This chapter contains the background of the study, research problem, objectives, and justifications of the study.

1.1 Background

Academic transition is affected by personal, structural, and academic factors. This is regardless of whether one is transiting from nursery school to primary, to secondary and finally to college. The personal factors make transition one big experience for an individual. This is because of the many changes that call for adjustments if one is to fit into the new system. Among the changes are: how to make new friends in unfamiliar group of colleagues, knowing new teachers, navigating through the school, the subject groupings, coping with the boarding school environment for the ones who came from day schools just to mention a few. Another thing is that students often have to deal with new perceived pressures on academic performance from teachers who are keen to help the student get into the next academic level, social support from parents, sense of school belonging and other stressors.

Another challenge that students face on their way up is peer pressure. Certain vices like engaging in drugs, alcohol and sex gains strength as an individual grows through the system. Although the vices are highly found in high schools and colleges, lower levels also have minor cases that cannot be ignored. However, every change has an effect on the academic transition of an individual in any setting.

Much of the research on factors affecting academic transition has revolved around transition to first grade, transition to high school and transition to higher education. There is no single research sampling out the most effective factors contributing to academic transition. This research will thus collect all the identified factors affecting academic transition from the available literature and use the principle component analysis (PCA) and multiple regression to identify the most effective factors that contribute to academic transition.

1.2 Problem Statement

The measurement of academic transition has been based on performance in the Kenyan education system. Every researcher who has considered a research in this field has identified factors that are either related to other research findings or are entirely different in wordings but same in effect to academic transition. Utilizing the already identified factors affecting academic transition in Kenya but reducing them to the most effective ones could lead to a better understanding of the underlying factors that influence the academic transition process in Kenya.

1.3 Research Objectives

1.3.1 Primary Objectives

To model the factors affecting transition from one academic level to the other using principal component analysis and multiple regression.

1.3.2 Secondary Objectives

1. To analyze the factors contributing to academic transition in Kenya using multivariate analysis - principle component analysis.
2. To determine the most effective factor that contributes to academic transition.

1.3.3 Research Questions

i. Is the data set appropriate for describing the problem using multivariate and principle component analysis?

ii. Does multivariate analysis using principle component analysis help to identify the factors contributing to academic transition in Kenya?

iii. What are the best combinations of factors that could contribute to academic transition?

1.4 Justification and Significance of the Study

The approach taken in this project will summarize other works done on effects to academic transition. This modeling technique aims at creating the best model that captures as many factors as possible. Furthermore, modeling academic transition using principal component analysis (PCA) will create ranks on the factors hence be used to advice education policy makers on the areas that need more emphasis for an effective transition program since this model can be used reliably in setting primacies in areas of interest in an education program.

1.5 Outline

The subsequent sections of this Project are organized as follows: chapter two, literature review on factors affecting academic transition; chapter three, describes the study methodology; chapter four provides the results and data analysis; chapter five is on the study conclusion and recommendations.

CHAPTER TWO: LITERATURE REVIEW

2.0 Introduction

This chapter explored literature on academic transition with an aim of mining as many identified factors as possible. The identified factors will be used as variables in principle component analysis later in data analysis.

2.1 First Grade Transition

Entwisle and Alexander (1998) conducted a research on the nature of transition to first grade and factors affecting transition. Using a longitudinal study of children who joined first grade in Baltimore in 1982, the study focused on factors affecting transition based on type of schooling (full-day versus half day kindergarten), and type of family arrangement(nuclear versus extended family) (Entwisle and Alexander, 1998). The test scores on the multivariate models fitted indicated that family structures had a bigger influence on transition in that children from single parent families do not do well as compared to those with both parents. Early entry into the system affects children IQ, which consequently affects performance. Kindergarten experience and academic performance also affects how well the child adopts to the new learning environment. For example, children who understood sounds and performed exceptionally well in kindergarten have less difficulties in achieving good grades in the first grade.

2.2 Primary School Transition to Secondary School

Research division of Ministry of Education New Zealand (2008) conducted a study on transitions of students from primary school to secondary school. The study involved following 100 students over 18 months as they transitioned from primary school all the way to year 10. One of the key purposes of the study was to identify the factors that affect smooth transition of students from primary school to high school. Using Assessment Tools for Teaching and Learning (asTTle), the division found that academic transition is highly dependent on student attitude and engagement in learning, perception of the school and teachers by students, class grade since year 9, which is more challenging than the rest, home circumstance, and learning environment in terms of noise or disruptions.

Evangelou, et al., (2008), conducted a sub-study on the factors that make an effective transition from pre-school, to primary and secondary education. Using a mixed method approach, the study identified that among the social- cultural factors affecting transition were ethnicity where children from minority ethnic backgrounds had negative transition processes. Language spoken at home other than English also had a negative effect on transition and religion affected the student integration in the classroom thus affecting the success of the classroom. The above social-cultural factors affects the child's self-esteem and ability to make new friends, amount of time needed for the child to settle down in a new school, pace of adjusting to routines and organization of a new school environment, developing interest in school work, and the ability to understand and connect curriculum to previous level.

2.3 Parental Impact on Successful Academic Transition

Baker and Stevenson (1986) did a research on how mothers assist children in managing transition to high school. By expanding the extant model fitted on data collected from an exploratory study of a heterogeneous sample of mothers of eighth graders transiting to high school, the researcher explored the relationship between social-economic status and academic performance. Of the status considered were parent's level of education, parent's occupation, family income, and race. By use of correlation, the researchers were able to find out that parents especially mother's play a big role in influencing a child's career directly. Their level of influence is highly based on the parent's level of education. The strategy a mother employs is highly dependent on her understanding of the child and the curriculum. Mothers with a higher level of education tend to assist their children more since they are able to follow up their children's performance as well as keep contact with the school management to be able to monitor the child's behavior and performance. The research also showed that whites performed better than blacks in transition to high school.

2.4 Social, Personal, and Academic Factors Affecting Transition

Jimmetra (2010) examined the structural, personal, and academic factors affecting transition of Africa Americana males from primary to high school levels. Using explanatory mixed model design, data on eight themes related to physical structure, academic structure, teacher expectations, and academic expectation. In additional to this were: academic assistance, relationship with peer, teacher student relationship, home and community relationship, and social psychological issues, data was collected from 16 students who had just completed first semester of freshmen year (Jimmetra, 2010). Descriptive and inferential statistics was used. Quantitative data was collected while Creswell's content analysis coding procedure was used for qualitative data (Jimmetra, 2010). The results identified that of the eight themes teacher expectation, academic expectation, relationship with peers, home, and community relationship affected transition. Structural and academic structures and teacher expectation though not highly considered they might cause some struggle especially when the school is overcrowded. Academic factors like time management, curriculum alignment, attendance academic commitment, and grades have a direct effect on academic achievement and transition.

Warunga et al., (2011) undertook a research on factors affecting education transition rates in Taita-Taveta district, Kenya. By considering transition rates by gender, the factors related to culture, environment, school, and social economic status was seen to affect academic transition in a big way. From the literature, the following factors were identified: school regulations and management style, school curriculum, student attitudes, teachers, physical facilities, parent's occupation and level of education, family size, parental engagement, birth order, student's career ambitions, and gender of the student. From a sample of 144 respondents where, 88 were parents and 56 primary school head teachers, the following factors were identified: lack of money, early marriages, school proximity, peer influence, lack of interest in schooling, drug abuse, domestic chores, parental ignorance, acquired a job and physical disability. University educated parents had a higher opinion on enrollment of children to secondary school as compared to those with lower education levels. Of the children who failed to join form one, Warunga et al (2011) identified that they engage in social economic activities like domestic chores, casuals in sisal plantations, hawking, farming, casual hire, herding and fetching firewood in level of engagement.

2.5 Physical Development Effect on Academic Performance

Isakson and Jarvis (1998) did a short-term longitudinal study on the adjustments of adolescents during the transition into high school to eighth graders attending schools connected to public university. Issues like sense of independence, professed stressors, social support, and school ownership of the student, performance, and school attendance were put on check on all stages of the study. Through the study, parent's participation, which depends on the level of education, was paramount in facilitating a smooth adjustment to the transition. Using descriptive statistics, school membership, adaptive coping, sense of autonomy, daily hustles had the highest influence measured by their means in descending order. The rest of the factors like grade point average, school attendance friends, and parents had least effect on adjustment. However, parental and friends support had a significant relationship to school membership since any relationship changes affects how a child felt about the learning and home environment thus influencing their transition to the next grade.

2.6 Gender Impact on Schooling Transition

Dube (2011) conducted a research on "factors affecting transition, performance, and retention of Girl's in secondary school in arid and semi-arid land in Mandera county of Kenya". The study involved an in-depth literature review on factors affecting transition. Among many factors identified from the literature, the following were outstanding. Gender differences where males had access to education more than girls, place of residence either urban or rural, harsh weather conditions, infrastructure, security and teacher availability, economic activity of the region, selection criteria to high school and college entry, social status like poverty, household size, insufficient school supply and high cost of fees. Using both qualitative and quantitative data, from a sample population of 1280 respondents, descriptive statistics identified the following factors. Historical factors, social cultural factors like early marriages, social-economic factors like access to school facility, and environmental factors like school structure, availability of learning materials, congestion and water and sanitation, also affect transition, entry, and retention of the girl children from primary schools to secondary schools. Using both qualitative and quantitative analysis techniques, Dube (2011) was able to identify parent's level of education, marital status, and age distribution of the parents as factors that contributed to academic transition of the child. Among the many factors, lack of school fees affects transition from primary school to secondary despite the fact that Kenya secondary education was declared free in 2008.

Using different methods, every researcher on education transition has identified a number of factors that could affect academic transition. However, none of them has grouped the factors in levels of influence. Therefore, these factors can be grouped as school and student factors, demographic factors, parents and students factors, curriculum and school factors, social psychological factors, social cultural factors, environmental and security factors and social economic factors. The following conceptual framework is a summary of all the factors in those categories.

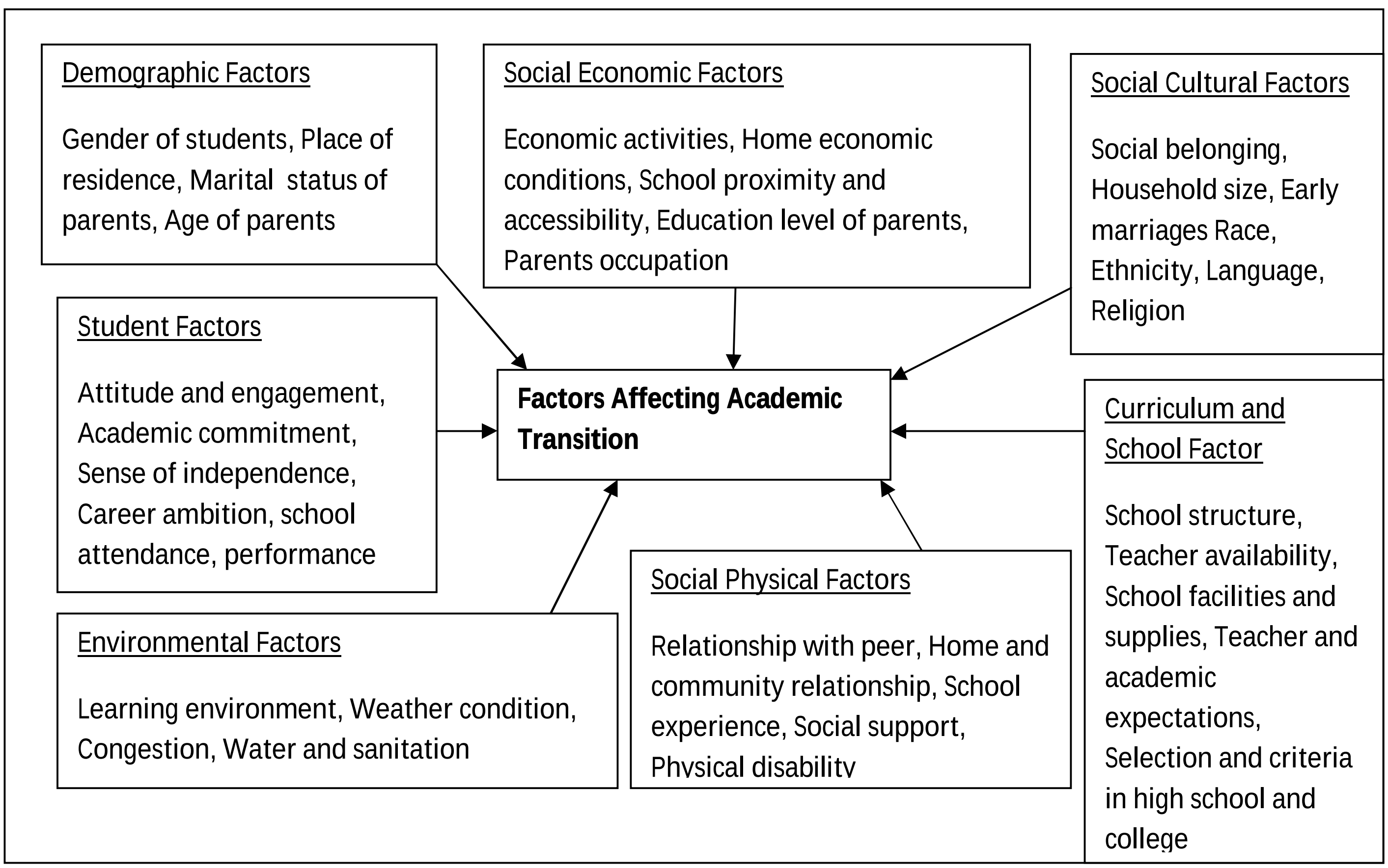

Figure 2.1: Framework on Factors Affecting Academic Transition

CHAPTER THREE: METHODOLOGY

This chapter explains the model to be used. The paper will investigate the factors affecting academic transition by use of Multivariate analysis - Principle component analysis (PCA) and multiple linear regression.

3.1.Research Design

The study adopts Kenya demographic health survey (KDHS) survey design. KDHS used descriptive survey design since it is ideal for collecting information on human behavior, opinion, and feelings on issues surrounding health and education. The survey method is appropriate for collecting data aimed at evaluation and decision-making. It is thus most appropriate for collecting data that evaluate social systems in the society.

3.2.Study population

The target population for this study was all women in the eight main regions of Kenya. The regions were inclusive: Nairobi, Central, Eastern, Western, Nyanza, North Eastern, Coast, and Rift Valley. Sample populations of 6079 women were interviewed in particular KDHS study.

3.3.Model Formulation

This part explains the two models to be used in this project. These are multivariate analysis models and specifically principal component analysis and multiple linear regressions.

3.3.1 Multivariate analysis

Multivariate data analysis is concerned with techniques of analyzing continuous quantitative measurements on several random variables. It forms a systematic study of *m* random variables

$$X' = [x_1 , x_2 , \dots \dots \dots x_m] \tag{3.1}$$

The mean value of the random *m* x 1 vector is a vector of means

$$E(\mathrm{X}) = \begin{pmatrix} Ex_1 \\ \cdot \\ \vdots \\ Ex_m \end{pmatrix} = \int x\, f(x)dx = \begin{pmatrix} \int x_1 f(x)dx \\ \cdot \\ \vdots \\ \int x_m f(x)dx \end{pmatrix} = \mu$$

$$(E\ (X))' = [\ E\ (X_1), E\ (X_2), \dots \dots \dots . . E\ (X_m)] \tag{3.2a}$$

$$\mu = E\ (X) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_m \end{bmatrix} \tag{3.2b}$$

Covariance is a measure of dependency between two random variables X and Y (Hardlean Simar, 2003). The theoretical covariance of a:

X ~ (μ_X, Σ_{XX}) and Y ~ (μ_Y, Σ_{YY}) is a (m x q) matrix of the form

$$\Sigma_{XY} = cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)^T$$

The covariance matrix of a random vector X is an m *x m* matrix defined by

$$cov\ (X) = E\ \{[X - E(X)][X - E(X)]'\} \tag{3.3}$$

$$= E\{[X - \mu_X][X - \mu_X]'\}$$

$$\Sigma_{XX} = \begin{bmatrix} \sigma_{11}\sigma_{12} \cdots & \sigma_{1m} \\ \sigma_{21}\sigma_{22} \cdots & \sigma_{2m} \\ . \quad . & . \\ . \quad . & . \\ . \quad . & . \\ \sigma_{m1}\sigma_{m2} \cdots & \sigma_{mm} \end{bmatrix} = \Sigma$$

Where

$$\sigma_{ij} = cov\,(x_i\ , x_j) = E\{[x_i - \mu_i][x_j - \mu_j]\}$$

And

$$\sigma_{ii} = \sigma_i^2 = E[(x_i - \mu_i)^2] = var\,(x_i)$$

The vector X with μ_X and covariance matrix Σ is written as

$$X \sim (\mu_X, \Sigma_{XX})$$

It follows that $\Sigma_{XY} = \Sigma_{Y^T X}$ and $Z = \binom{X}{Y}$ has a covariance matrix

$$\Sigma_{ZZ} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY} & \Sigma_{YY} \end{pmatrix}$$

From $\sigma_{xy} = cov(X, Y) = E(XY^T) - (EX)(EY)^T = E(XY^T) - \mu_X \mu_Y{}^T$

It follows that $cov(X, Y) = 0$ when X and Y are independent and $cov(X, Y) = 1$ when dependent. This follows that the diagonal elements of Σ_{ZZ} must be non-negative and the Σ_{ZZ} matrix should be positive definite matrix (Timm, 2002).

Eigenvalue and Eigenvectors**:** For every square matrix A, there exist a scalar λ. A nonzero vector x can be found as

$$A\underline{x} = \lambda\underline{x} \qquad (3.4)$$

This can be written as

$$(A - \lambda I)\underline{x} = 0$$

where λ is called an Eigenvalue of A, and $\underline{x}$ is the Eigen vector of A corresponding to λ (Rencher, 2002). Therefore, if A is an m x m matrix, the characteristic equation will have m Eigen values *i.e.* $\lambda_1, \lambda_2, \ldots \ldots \ldots \ldots \lambda_m$ and corresponding Eigen vectors $x_1, x_2, \ldots \ldots \ldots \ldots x_m$.

3.3.2 Demonstration of Variable Correlation

This part contains a fictitious example related to the topic of the study. Imagine you have six questions to be used as measures of successful transition. Each of these questions is to be used as an element of a set of six predictor variables in a multiple regression equation where the dependent variable is ability to transit to the next academic level. An empirical finding assists in identifying if there exists correlation or not. Obtaining a correlation matrix assists in making conclusions on the state of correlation among the variables.

Table 3.1: Demonstration Of Correlation Matrix

Correlation

Variables	1	2	3	4	5	6
1	1.000					
2	.288	1.000				
3	-.753	-.686	1.000			
4	.103	.971	-.624	1.000		.
5	.854	-.052	-.672	-.158	1.000	
6	.449	.964	-.769	.905	.115	1.000

3.3.3 How to Interpret a Correlation Matrix

The rows and the columns of Table 3.1 correspond to the six variables considered in the analysis. Row 1 and column 1 represents variable 1, row 2 and column 2 represents variable 2 and so forth. The intersection of a given row and column marks the correlation between two corresponding variables. For example, the correlation between variable 3 and variable 5 is the intersection between column 3 and row 5 which is -0.672. The strength of the correlation is measured on a 0 to 1 scale. The closer the correlation to 0, the lower the level of correlation, the closer it is to 1 the higher the level of correlation. Thus, in the example above, variables 2 and 4, 2 and 6, and 4 and 6 have higher correlations, 1and 4, and 5 and 6 have lower correlations.

The level of correlations either high or low thus orders the variable groupings in PCA. The variables with higher correlation are assumed to be measuring the same construct thus can be grouped together. For example, the highly correlated variables 2, 4 and 6 could be grouped as measuring “availability of resources” and the variables 1, 5 and 6 could be measuring “gender effect on transition”. The reduction process in PCA thus picks the number of construct being measured as a way of removing redundancy in the data.

3.4 Principal Component Analysis

What is a Principle Component? According to Hatcher and O'Rourke (2013), a principal component is a linear combination of optimally weighted observed variables. The weight is well described by the subject scores on a given principal component computed. For example in the above fictitious study, each subject would have scores on the two components, one score on availability of resources and one score on gender effect on transition. The subject’s actual scores on the six questions would be weighted and their sum will compute their scores on a given component. The scores on the first component extracted can generally be computed as

$$y_1 = e_{12}x_2 + \cdots\cdots + e_{1j}x_j \tag{3.5}$$

y_1 the subject’s score on the first principle component ,e_{1j} the regression coefficient for the observed variable j ,x_j the subject scores on the observed variable j

3.4.1 Theory of Principal Component Analysis (PCA)

Principal component analysis (PCA) is used when one needs to extract a small number of artificial $y_1, y_2, \dots, y_q$ variables from a large number of observed variables $x_1, x_2, \dots, x_q$. The artificial variables are thus called the principal components. The first few principle components preserve the highest level of variations present in the original set of variables. The extracted artificial variables are used as predictor variables in the succeeding analysis.

3.4.*2* Algebraic Basis of Principal Component Analysis

i. The first component y_1 is a linear combination of the optimally weighted observed variable illustrated in Equation 3.6, which accounts for the highest amount of the total variation in the observed variables i.e. it must have correlated with many of the observed variables.

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots .+a_{1q}x_q = \underline{a'_1x} \tag{3.6}$$

To restrict the exponential growth of the variance of y_1 a restriction$a_1'a_1 = 1$ is applied.

ii. The second component y_2 is a linear combination of the observed variables that account for the greatest amount of variance in the observed variables that was not accounted for by y_1. The second component is given in Equation 3.7.

$$y_2 = a_{21}x_1 + a_{22}x_2 + \dots .+a_{2q}x_q = \underline{a'_{21}x} \tag{3.7}$$

The variance is subject to the following two conditions:

$$a_2'a_2 = 1$$

The first component y_1 is uncorrelated with the second component y_2.

$$a_2'a_1 = 0$$

iii. The q^{th}principal component is a linear combination of observed variables that account for the greatest amount variance that was not accounted for by the preceding components given the following conditions:

$$y_q = a_{q1}x_1 + a_{q2}x_2 + \dots .+a_{qq}x_q = \underline{a_q'x} \tag{3.8}$$

$$a_q'a_q = 1$$

$$a_i'a_q = 0\ (i \le q)$$

$a_{1,}, a_2, \dots . a_q$ Satisfies this condition corresponding to the eigenvalues $\lambda_{1,}\lambda_2 \dots \dots \dots \dots \lambda_q$ of the covariance matrix S

If the eigenvalues of S are $\lambda_{1,}\lambda_2 \dots \dots \dots \dots \lambda_q$, and since $a_q' a_q = 1$, then the variance of the qth principal component is λ_q. The total variance of the q principal components is equal to the total variance of the original variables.

If $S = \{s_{ip}\}$ is the $p \times p$ sample covariance matrix with eigenvalue, eigenvector pairs(λ_1, e_1), (λ_2, e_2) , …, (λ_p, e_p), the pth sample principal component is given in Equation 3.9;

$$y_i = e_1 \underline{x} = e_{i1}x_1 + e_{i2}x_2 + \cdots + e_{ip}x_p \quad , i = 1,2, \dots , p. \tag{3.9}$$

Where $\lambda_1 > \lambda_2 > \cdots \dots \dots \dots > \lambda_p > 0$ and x is any observed variable $x_1, x_2, \dots , x_p$.

3.4.3 Difference between PCA and Exploratory Factor Analysis

More often than not, PCA has been mistaken to be equivalent to exploratory factor analysis due to the numerous similarities that exist between the two extraction methods. Among the similarities are both variable reduction procedures used to classify observed variables into groups, both procedures may yield the same results. However, the two have a big conceptual difference that needs to be made clear. Factor analysis assumes that presence of at least one latent variable cause co-variation in the observed variables hence exerting causal influence on the observed variables (Hatcher and O’Rourke, 2013). Unlike factor analysis, PCA does not make any assumptions on causal model; it simply reduces the variables into smaller number of components accounting for the highest variation in a group of observed variables (Hatcher and O’Rourke, 2013).

3.4.4 Calculating Principal Components

Suppose n independent observations are taken on $X_1, X_2, \cdots \cdots , X_p$, where the covariance between $X_i$and X_j is given in Equation 3.10

$$Cov\left(X_i, X_j\right) = \sum = R \tag{3.10}$$

$$i = 1, 2,3, \cdots\cdots\cdots, p$$

$$j = 1, 2,3, \cdots\cdots, p$$

Let $\lambda_1 > \lambda_2 > \cdots \ldots \ldots \ldots > \lambda_p > 0$ be the eigenvalues of the correlation matrix of the variables above and let the corresponding eigenvectors be $e_1, e_2, \ldots \ldots \ldots \ldots x_p$. Normalize the eigenvectors so that $e_p^T e_p = 1$.

Define y_1 to be the first principal component. The component is a linear combination of the X's which has the largest possible variance:

$$y_1 = \mathrm{a}_1^{\mathrm{T}} X = \sum_{i=1}^{p} a_{1i} X_i \tag{3.11}$$

$$a_1^T a_1 = 1$$

$$Var(y_1) = a_1^T \Sigma a_1 = \sigma^2 a_1^T R a_1$$

Constrained maximization leads to the Lagrangian:

$$L = a_1^T R a_1 + \lambda(a_1^T a_1 - 1)$$

Solution: $a_1$is a unit vector satisfying

$$R a_1 = -\lambda a_1$$

I.e. a_1 must be an eigenvector of R.

The objective was to maximize $a_1^T R a_1 = \lambda a_1^T a_1$ where λ is the eigenvalue corresponding the eigenvector a_1.

Therefore, take $a_1 = e_1$

That is

$$y_1 = e_1^T X = \sum_{i=1}^{p} e_{1i} X_i \tag{3.12}$$

y_1 has the largest variance among all linear combinations of the X's.

To calculate the second principal component**:** Let $y_2$be the second linear combination of the X's with the largest possible variance:

$$y_2 = a_2^T X = \sum_{i=1}^{p} a_{2i} X_i \tag{3.13a}$$

But Cor(y_1, y_2)=0 i.e.$a_2^T e_1 = 0$

The constrained maximization leads to the Lagrangian

$$L = a_2^T R a_2 + \lambda_1(a_2^T a_2 - 1) + \lambda_2(a_2^T x_1)$$

Solution:

$$(R + \lambda_1 I)a_2 + \lambda_2 x_1 = 0$$

With $a_2^T a_2 = 1$, $Re_1 = \lambda_1 e_1$ and $a_2^T e_1 = 0$.

Multiply through by $\mathrm{e}_1^{\mathrm{T}}$ to find that $\lambda_2 = 0$

$$Ra_2 = -\lambda_1 a_2$$

Therefore,a_2 is an eigenvector (which cannot be X_1because of the orthogonality condition).

$a_2 = e_2$ will make $\mathrm{Var}(y_2) = \sigma^2\lambda_2$ which as large as possible under the given constraints.

That is

$$y_2 = e_2^T X = \sum_{i=1}^{p} e_{2i} X_i \tag{3.13b}$$

y_2 has the largest variance among all linear combinations of the X's which are orthogonal with y_1.

To calculate other principal components**:** y_j is the linear combination of the constraint that y_j is uncorrelated with $y_1, y_2, \dots \dots \dots, y_{j-1}$.

The constrained maximization problem is solved by setting

$$y_j = e_j^T X = \sum_{i=1}^{p} e_{ji} X_i \tag{3.14}$$

The variance of y_j is $\sigma^2 \lambda_j$

3.4.5 Principal Component Analysis Procedure

As expressed in the Equation 3.12, PCA is a stepwise procedure involving decisions making at every step of the procedure.

Step 1: Initial Extraction

This step involves extracting a number of principal components, which are equivalent to the number of variables involved in the analysis. The first component accounts for the highest total variance while the succeeding components will account for lower amount of variance progressively. This procedure extracts a large number of components. However, only the meaningful components in terms of total variation are needed for the analysis. Eigen values represent the total of variation accounted for by a particular component.

Step 2: Determining the Number of Principal Components to Extract

In PCA, the statistician decides the number of component to retain for rotation and interpretation. The expectation is that the first few will explain an important variation (Rencher, 2002). The choice of the principle components to retain is based on the following criteria:

a. Eigenvalue one Criterion/ Kaiser Criterion

This criterion holds and interprets any component with an Eigen value greater than one. This is because each observed variable contributes one unit of variation to the total variation. Therefore, a component with an eigenvalue greater than one will account for a greater variance than would be accounted for by one variable. Components with eigenvalue less than one are considered trivial and thus eliminated from the model for the sake of variable reduction.

The eigenvalue method is simple, not subjective since a component is retained simply because its eigenvalue is greater than one. When variable communalities are high for a moderate number of variables, the criterion is most efficient. The disadvantage of this method is that, when communalities are small for a large data set, the criterion can retain the wrong number of components. The components with eigenvalues closer to one e.g. 0.99 are considered meaningless and yet the difference between such a value and one is trivial.

b. Scree Test

This is a plot of eigenvalues associated with each component. The breaking point separates the components with large eigenvalues from those with small eigenvalues. The components that fall before the break are retained while those that come after the break are considered less important thus discarded. In cases where the scree plot displays more than one large break, the last large break is considered. The components before this large break are considered and the rest are discarded. The scree plot method provides reasonably accurate results when the sample is over 200 and the variable communalities are large. Despite being useful in most cases, it is not easy to tell the scree plot break point especially in social science research. Such ambiguity calls for further criterion like eigenvalue one criterion.

c. Proportion of Variance Accounted For in the model

This criterion involves retaining a component if it accounts for a set proportion. The proportion is obtained by

$$\text{Proportion} = \frac{Eigenvalue\ for\ \text{the}\ component\ of\ interest}{Total\ Eigenvalues\ for\ \text{the}\ correlation\ matrix}$$

A specified proportion e.g. 5%, 10% of the total variation differentiates the components to be retained. The advantage of the proportion accounted for is that it allows one to retain as many components as one wishes.

Step 3: Rotation

A rotation is a linear transformation formed on the factor solution for making the solution easier to interpret. It involves reviewing the correlations between the variables and the components for using the information to interpret the components. Ideally rotation helps in determining what contrast seems to be measured by component 1, 2 and so forth. The best rotation is varimax rotation, which is an orthogonal rotation that results in uncorrelated components. Compared to some other type of rotations a varimax rotation tends to maximize the variance of a column of the factor pattern matrix.

Step 4: Interpreting the Rotated Solution

This involves determining what each of the retained component measure, which involves identifying the variables that demonstrate high loadings for the principal component. Usually a brief name is assigned to each retained component that describes its content. The first decision to be made at this stage is to decide how large a factor loading must be to be considered large. The rule of thumb advice a loading of above 0.3 is large enough to be retained. The rotated factor pattern is interpreted as follows:

1. Read across the row for the first variable and drop the variables that load on more than one component. This is because such variables are not pure measures of any one construct.
2. Repeat this process for the remaining variables, scratching out any variable that loads on more than one component.
3. Review all the surviving variables with high loadings on component 1 to determine the nature of this component
4. Repeat this process to name the remaining retained components
5. Determine whether this final solution satisfies the following interpretability criteria:
    a) Each component should have at least three significant loadings retained
    b) The variables loading to a given component must share conceptual meaning
    c) Different variable loading on different components must be measuring different constructs

d) The factor pattern must demonstrate a simple structure.

Step 5: Creating Component Score

A component score is a linear combination of the optimally weighted observed variables. It indicates where the subject stands on the retained components. With this done, this component, scores could be used either as predictor variables or criterion variables in subsequent analysis. Remember that a separate equation with different weights is developed for each of the retained components.

Step 6 Applying the Chosen Principal Components to Regression

This step involves regressing the chosen principal components against the response variable. The covariance between PC vector y_j and the original vector X:

$$Cor(X, Y) = \frac{Cov(X,Y)}{\sqrt{var(X)var(Y)}} \quad (3.14)$$

Correlation between variable X_i and the principal component $y_j$may be seen as a proportion of variance of the i^{th} variable explained by the j^{th} principal component y_j. The correlation between the i^{th} and the X_i and the j^{th} is the factor score (loading) of the i^{th} variable X_i with the j^{th} principal component given in Equation 3.15:

$$F_{ij} = \frac{V_{ij} \times \sqrt{\lambda_j}}{S_i} \quad (3.15)$$

$F_{ij} - factor\ loading\ for\ the\ i^{th}\ variable\ and\ j^{th} principal\ component\ factor$

V_{ij} -normalized eigenvector value for the $i^{th}\ variable\ in\ j^{th}\ eigen\ vector$

λ_j - j^{th} Eigenvalue

S_i – Standard deviation of the i^{th} $variable$

RegressY = $\beta_0 + \beta_1 y_1 + \beta_2 y_2 + \cdots\cdots + \beta_q y_q + \varepsilon$

Instead of Y = $\beta_0 + \sum_{i=1}^{m} \beta_i y_i + \varepsilon$

$$q < m$$

y_1, y_2 and y_3 are orthogonal so t-tests for coefficients are easy to interpret.

Later, the identified components will be regressed with the level of the highest education.

3.5 Multiple Linear Regression Models

This technique allows additional variables to enter the analysis separately so that the effect of each can be estimated on the independent variable. It is valuable for quantifying the impact of various simultaneous influences upon a single dependent variable.

A linear model relating to the response variable *y* to several predictors has the form

$$Y = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \cdots + \beta_q y_q + \varepsilon \qquad (3.16)$$

The parametersβ_0, $\beta_1, \dots, \beta_q$ are the regression coefficients $y_1, y_2, \dots, y_q$are the model predictors (principle components). $\mathcal{E}$ is the error term providing information on the random variation in Y not explained by the *x* variables. The variation could be due to other variables not effectively included in the model. The multiple regression models in this project will be used to summarize the effect of each component to the education level.

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

4.1 Variables Description

The identified factors affecting education transition are summarized in Table 4.1**.**However, only a number of them will be used in the model due to the limitation of the data available. These are:

Y_1 = Highest educational level, X_1= Literacy, X_2 **=** De facto place of residence, X_3= Region, X_4= Religion, X_5= Ethnicity, X_6 = Type of place of residence, X_7= Number of household members, X_8= Index to birth history, X_9= Relationship to household head, X_{10}= Sex of household head, X_{11}= Has electricity, X_{12}= Main floor material, X_{13}= Main wall material, X_{14} = Main roof material, X_{15}= Water and Sanitation, X_{16}= Access to information, X_{17}= Mode of transport, X_{18}= Preventive health measures, X_{19} = Type of cooking fuel, X_{20}= Wealth index factor score, X_{21}= Parents marital status, X_{22}= Number of other wives, X_{23}= Wife rank number, X_{24}= Partner's education level, X_{25}= Highest year of education, X_{26}= Partner's occupation, X_{27} = Mother's occupation, X_{28}= Mother Work for family, others, self, X_{29}= Mother Works at home or away, X_{30}= Who decided how to spend money and X_{31}= Mother's type of earnings for work

Data analysis was conducted using SPSS version 20.1 package. By performing PCA on the 31 variables in SPSS, the correlation matrix showed that all the 31 variables were correlated (See Appendix 1). Thus, there was no need for eliminating any of the variables in the analysis at this point.

4.2 Principal Component Extraction using the Eigen Value Criterion and scree plot methods

The eigenvalue for the 31 variables where λ_1= 5.488, λ_2=2.607, λ_3=2.174, λ_4=1.899, λ_5=1.734, λ_6=1.477, λ_7=1.416, λ_8=1.285, λ_9=1.183, λ_{10}=1.097, λ_{11}=1.062, with λ_{12} to λ_{31} being less than one (see Appendix 2, column 2).

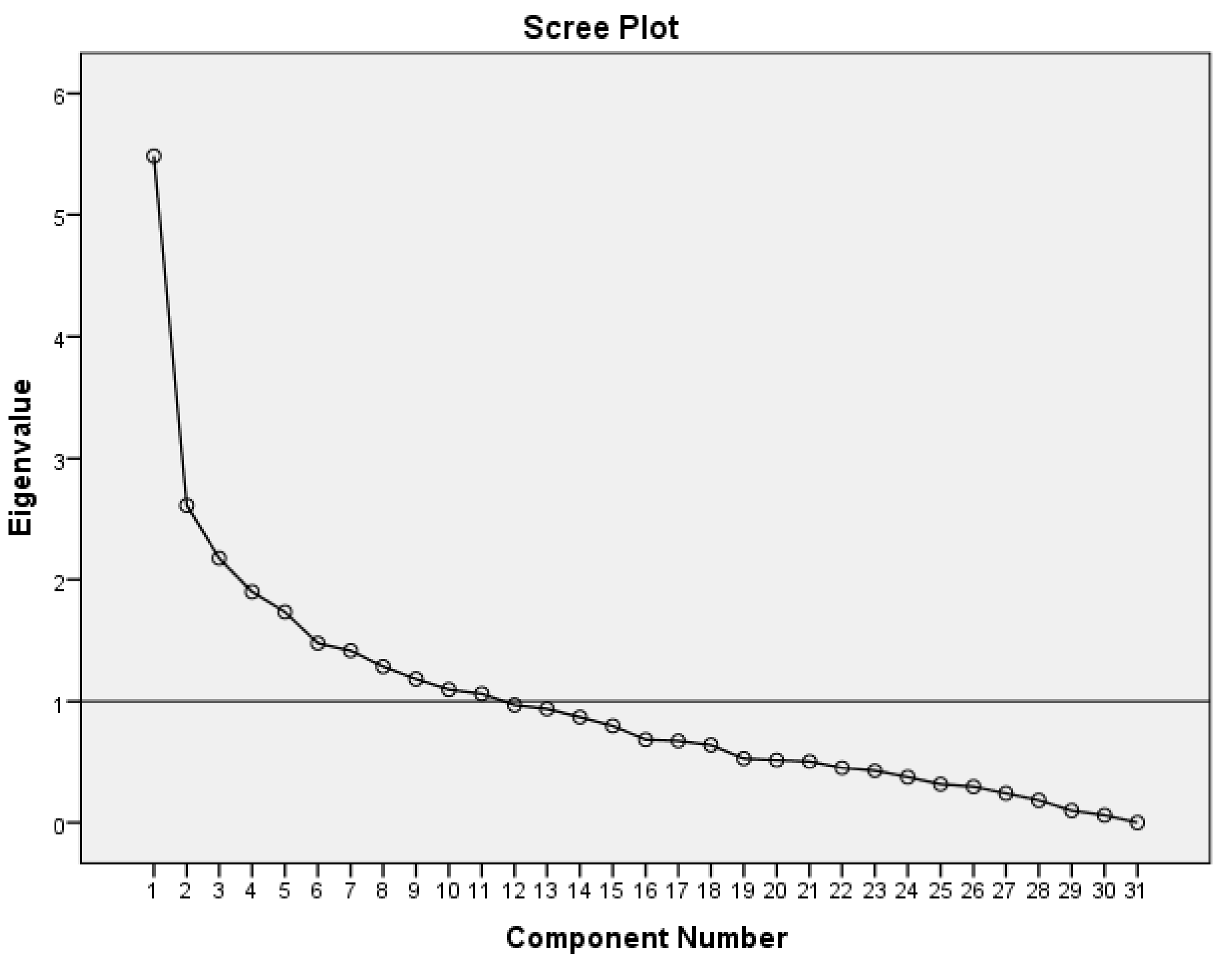

Figure 4.1: A Scree Plot of the Main Principle Components

Using the eigenvalue 1 mark on the scree plot, only 11 components have eigenvalues greater than 1, the circled points about the mark point (Figure 4.1).

Proportion of variance accounted for by each PC:

$$\text{Proportion} = \frac{Eigenvalue\ for\ \text{the}\ component\ of\ interest}{Total\ Eigenvalues\ for\ \text{the}\ correlation\ matrix}$$

$$\sum_{i=1}^{m} \lambda_i = \text{trace(S)}$$

$$p_m = \frac{\sum_{i=1}^{q} \lambda_i}{trace(S)}$$ p is the proportion of variance accounted for by q principal components

With a threshold of 70%, the components with meaningful cumulative proportions are component 1 to 11(See appendix 3, column 4).

Table 4.1: The Component Score

Component Score Coefficient Matrix

	Component										
	1	2	3	4	5	6	7	8	9	10	11
Literacy	.035	.091	.081	-.009	.356	.057	.139	-.009	-.160	.053	-.144
De facto place of residence	.216	.053	.068	-.048	.030	.030	.064	.018	.048	.095	.064
Region	.055	-.003	-.107	-.050	.355	-.051	-.123	.129	.052	.213	.164
Religion	.039	.080	-.020	-.138	-.483	.056	.116	.058	.026	.218	.019
Ethnicity	-.020	-.024	-.024	.022	.021	-.018	.011	-.002	.096	.019	-.675
Type of place of residence	.214	.023	.079	-.009	.013	.050	.160	.033	.034	-.069	.027
Number of household members	.020	-.130	.126	.296	-.200	-.062	-.010	-.049	-.181	.039	.088
Birth Index	.011	-.032	-.001	.115	.084	.072	-.246	.077	-.300	-.043	.115
Relationship to household head	-.010	.065	.035	.418	.009	-.024	.041	-.024	.007	.071	-.039
Sex of household head	.036	.010	-.007	-.439	-.135	.028	-.011	-.042	-.043	-.019	.042
Had electricity	-.100	-.041	-.028	.012	.004	.104	.095	.145	.046	-.210	-.054
Main floor material	-.151	-.002	.012	-.035	.024	-.063	-.032	-.061	-.094	.135	.122
Main wall material	-.031	.029	.382	.076	-.083	-.109	-.014	-.086	-.038	.189	.137
Main roof material	.055	.021	.417	.000	.064	.012	-.008	-.044	.009	-.023	.130
Had Water and Sanitation	.104	-.058	-.041	-.041	-.013	-.013	-.019	-.104	-.192	-.079	-.071
Access to information	-.066	-.021	-.088	.042	.085	.083	.234	-.081	.047	.073	.223
Mode of transport	.021	.001	-.035	-.001	.007	.388	.018	-.117	.121	-.202	.197
Preventive health measures	.004	-.053	.011	.064	-.064	-.051	.094	.026	.009	.603	-.026
Type of cooking fuel	.190	.063	.061	-.043	.038	.007	-.029	.060	.033	.206	.176
Wealth index factor score (5 decimals)	-.185	-.019	-.011	-.025	.030	-.008	-.007	-.024	.030	.032	.062
Parents marital status	.061	-.039	.384	.030	-.039	.028	-.045	.116	.189	-.103	-.215
Number of other wives	.029	.461	.008	.000	.011	.032	-.015	-.021	-.006	.004	.023
Wife rank number	.028	.462	.007	-.001	.011	.033	-.016	-.021	-.007	.005	.024
Partner's education level	-.012	-.017	.040	-.029	-.030	.005	.383	.044	-.085	.098	-.087
Partner's Highest year of education	-.078	.023	.024	-.078	.073	.063	-.513	.013	.061	.006	-.067
Partner's occupation	-.013	.027	.013	.049	.069	-.448	-.045	.011	.183	.023	.013
Mother's occupation	-.020	-.151	-.045	.010	-.072	-.398	.174	-.021	.028	-.285	.197
Mother Work for family, others, self	-.011	-.026	.048	.057	.013	.096	.018	-.556	.010	-.192	.028
Mother Works at home or away	.040	-.075	.068	.086	.048	-.011	.017	.434	-.161	-.158	.030
Who decided how to spend money	-.014	.000	-.078	.013	-.114	.235	-.177	.283	.359	.010	.094
Mother's type of earnings for work	.023	-.023	.117	.040	-.014	-.091	-.016	-.113	.579	-.002	-.091

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Extracted Principle Components

The extracted components were then explored in the following manner as shown below:

$$y_1 = 0.035x_1 + 0.216x_2 + 0.055x_3 + 0.039x_4 - 0.020x_5 + \cdots \ldots ..$$

$$y_2 = 0.091x_1 + 0.053x_2 - 0.003x_3 + 0.080x_4 - 0.024x_5 + \cdots \ldots ..$$

$$y_3 = 0.081x_1 + 0.068x_2 - 0.107x_3 - 0.020x_4 - 0.024x_5 + \cdots \ldots ..$$

$$y_4 = -0.009x_1 - 0.048x_2 - 0.050x_3 - 0.138x_4 + 0.022x_5 + \cdots \ldots ..$$

$$y_5 = 0.356x_1 + 0.030x_2 + 0.355x_3 - 0.483x_4 + 0.021x_5 + \cdots \ldots ..$$

$$y_6 = 0.057x_1 + 0.030x_2 - 0.051x_3 + 0.056x_4 - 0.018x_5 + \cdots \ldots ..$$

$$y_7 = 0.139x_1 + 0.064x_2 - 0.123x_3 + 0.116x_4 + 0.011x_5 + \cdots \ldots ..$$

$$y_8 = -0.009x_1 + 0.018x_2 + 0.129x_3 + 0.058x_4 - 0.002x_5 + \cdots \ldots ..$$

$$y_9 = -0.16x_1 + 0.048x_2 + 0.052x_3 + 0.026x_4 + 0.096x_5 + \cdots \ldots ..$$

$$y_{10} = 0.053x_1 + 0.095x_2 + 0.213x_3 + 0.218x_4 + 0.019x_5 + \cdots \ldots ..$$

$$y_{11} = -0.144x_1 + 0.064x_2 + 0.019x_3 - 0.675x_4 + 0.027x_5 + \cdots \ldots ..$$

Using the rule of thumb, retain all items with structure coefficients with an absolute value of .300 or greater (Table 4.2).

Table 4.2: Rotated Scores

Rotated Component Matrix[a]

	Component										
	1	2	3	4	5	6	7	8	9	10	11
Wealth index factor score (5 decimals)	-.906										
De facto place of residence	.890										
Type of place of residence	.865										
Type of cooking fuel	.773										
Main floor material	-.745										
Had Water and Sanitation	.636										
Had electricity	-.556										
Wife rank number		.959									
Number of other wives		.958									
Main roof material			.815								
Main wall material	-.419		.729								
Parents marital status			.709								
Sex of household head				-.850							
Relationship to household head				.835							
Number of household members				.557	-.398						
Religion					-.772						
Literacy					.626						
Region					.601						
Partner's occupation						-.734					
Mother's occupation						-.644				-.401	
Mode of transport						.605					
Partner's Highest year of education							-.721				
Partner's education level							.600				
Access to information	-.387						.419				
Birth Index							-.394		-.380		
Mother Work for family, others, self								-.787			
Mother Works at home or away								.654			
Mother's type of earnings for work									.719		
Who decided how to spend money						.350		.386	.458		
Preventive health measures										.746	
Ethnicity											-.811

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 14 iterations.

Table4.3: Table of Extracted Principal Components

y_1	**Social Economic status** as a factor of wealth index, de facto place of residence, type of cooking fuel, main floor materials, had water and sanitation, and had electricity.
y_2	**Family position** as a factor of wife rank number and number of other wives
y_3	**Home environment** as a factor of main floor materials, main wall materials, and parent's marital status
y_4	**Family composition** as a factor of relationship to household head, sex of household head, and number of household members
y_5	**Regional influence** as a factor of respondent's literacy level, religion, and region
y_6	**Parent's occupation** as a factor of partner's occupation, mother's occupation, and mode of transport
y_7	**Parent's education level** as a factor of partner's highest year of education, partner's education level, access to information, and respondent's birth index
y_8	**House wife status** as a factor of mother works for family and mother works at home
y_9	**Mother's type of earnings** as a function of mother's type of earning and who decides how to spend money
y_{10}	**preventive health measures**
y_{11}	**Ethnicity**

4.3 Regression of component scores with education level

This model tests the hypothesis

H_0: The components can be reliably used to describe change in academic levels (Transition)

H_1: The components cannot be reliably used to describe change in academic levels (Transition)

Table 4.4 Anova Table

Model		Sum of Squares	Degrees of freedom	Mean Square	F- Value	Sig.
	Regression	40.328	11	3.666	16.899	.000[b]
1	Residual	54.669	252	.217		
	Total	94.996	263			

Table 4.5 Regression Model Summary

Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
1	. 652[a]	.425	.399	.466

a. Predictors: (Constant), Ethnicity, Preventive Health Measures, Mothers Type Of Earnings, House Wife Status, Parents Education Level, Parents Occupation, Regional Influence, Family Composition, Home Environment, Family Position, Social Economic Status

The model summary in Table 4.65indicatesthat **R squared= 42.5%.**This implies that, the components account for only 42.5 % of change in education level. The rest of the percentage (51.5%) could be accounted for by other variables that were not considered for this project. Predictors: (Constant), Ethnicity, Preventive health measures, Mother's type of earnings, House wife status, Parent's education level, Parent's Occupation, Regional Influence, Family composition, Home environment, Family position and Social Economic status. With an α=0.05, and a calculated p-value=0.000, we can conclude that components can be reliably used to describe change in academic levels (Transition)

Table 4.6 Coefficient Estimates

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	95.0% confidence interval for B	
		B	Std. Error	Beta			**Lower bound**	**Upper bound**
1	(Constant)	1.004	0.029		35.017	0.000	.947	1.060
	Social Economic status	-.207	0.029	-.345	-7.210	0.000	-.264	-.151
	Family position	.070	0.029	.117	2.454	0.015	.014	.127
	Home environment	.047	0.029	.077	1.620	.106	-.010	.103
	Family composition	-.110	0.029	-.183	-3.824	.000	-.166	-.053
	Regional Influence	.220	0.029	.366	7.652	.000	.163	.276
	Parent's Occupation	.067	0.029	.111	2.316	.021	.010	.123
	Parent's education level	.133	0.029	.222	4.640	.000	.077	.190
	House wife status	.120	0.029	.199	4.170	.000	.063	.176
	Mother's type of earnings	-.020	0.029	-.033	-.683	.495	-.076	.037
	Preventive health measures	-.010	0.029	-.016	-.335	.738	-.066	.047
	Ethnicity,	.077	0.029	.128	2.687	.008	.021	.134

a. Dependant Variable: Highest Education Level

Using the coefficients in Table 4.6, the regression model can be written as

$$Y = 1.004 - 0.207y_1 + 0.07y_2 + 0.047y_3 - 0.11y_4 + 0.22y_5 + 0.067y_6 + 0.133y_7 + 0.12y_8 - 0.02y_9 - 0.01y_{10} + 0.077y_{11}$$

However, with an α=0.05 components $y_3, y_9, and\ y_{10}$ are not significant in the model, they all have a p-value greater than α. Therefore, these three components have little effect on academic transition. Never the less, the rest of the variables can be used to describe academic transition in levels.

Table 4.7: Multiple Regression Model Interpretation

Component	regression Coefficient	Interpretation
y_1	-0.207	Any change from a higher economic status to a lower economic status will reduce chances of transition by 0.207 times.
y_2	0.07	Any change from a polygamous family to a monogamous family will increase chances of academic transition by 0.07 times.
y_4	-0.011	Any change of relationship to family head will reduce academic transition by 0.011 times.
y_5	0.22	Any change from inability to read to ability to read will increase chances of transition by 0.22 times.
y_6	0.067	Parents' occupation affects transition by 0.067 times.
y_7	0.133	Any change on parents' education from no education to primary, to secondary, to higher education will increase chances of transition by 0.133 times.
y_8	0.12	Any change in mother's place of work increases chances of transition by 0.12 times.
y_{11}	0.077	Ethnicity plays a big role in transition. It affects transition by 0.077 times.

NOTE: All the above interpretation is based on situations when all the other variables are held constant, the effect of an individual principal component on transition.

From the above interpretation, the effect of each principal component ranked by the weight individual component places on transition is $y_5, y_1, y_7, y_8, y_4, y_{11}, y_2, y_6$.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

Education transition has been affected by many factors. This project sort to identify the most effective factors influencing transition. Among the identified factors from literature were given in Figure 2.1. Due to lack of enough data, this project used only 31 variables available to us. Out of the 31, 11 components were extracted (as given in Table 4.3). The most effective principal components to academic transition are regional influence, social economic status, parent's education, housewife status, family composition, ethnicity, family position and parents' occupation by level of effect. These components accounted for a total variation of 42.5% to education transition. This implies that, apart from the identified components, education transition is affected by many other variables not identified in this project. These variables could be similar to the ones highlighted in the factors affecting education transition framework (Figure2.1).

The recommendation I give to any researcher venturing into the area of academic transition in Kenya is that, they should consider the variables identified in Figure 2.1 but were left out in this project. The emphasis should be majorly on measuring curriculum and school factors, student factors, and social physical factors. The specifics around subjects like curriculum and school factors would be more relevant since the concern is on factors affecting academic transition and curriculum is believed to have a direct effect.

With regional influence taking the highest percentage of effect on transition, regional balance will be key when stakeholders consider issues on development. Enhancing accessibility of both the education facilities and facilitators will go a long way in reducing regional imbalance.

REFERENCES

Baker D. P. and Stevenson D. L., (1986): Mother's Strategy for Children's School Achievement: Managing the Transition to High School. *Sociology of Education,* Vol. 59.pp. 156-166.

Dube A. K. (2011): Factors Affecting Transition, Performance, and Retention of Girls in Secondary Schools in Arid and Semi-Arid Land: A Case of Rhamu Town-Mandera County, Kenya. Department of Educational Management, Policy and Curriculum Studies, 82P.

Entwisle D. R. and Alexander K. L., (1998): Facilitating the Transition to First Grade: The Nature of Transition and Research on Factors Affecting it.*The Elementary School Journal,* vol. 98.pp 351-364.

Evangelou, Maria and Taggart, Brenda and Sylva, Kathy and Melhuish, Edward and Sammons, Pam and Siraj-Blatchford, Iram (2008): *Effective Pre-school, Primary and Secondary Education 3-14 Project (EPPSE 3-14): What Makes a Successful Transition from Primary to Secondary School*?Institute of Education, University of London 2008.

Hardle W. and Simar L., (2003): *Applied Multivariate Statistical Analysis*. Methods of data Technologies.

Hatcher Larry and O'Rourke Norm.(2013).A Step-by-step Approach to Using SAS for Factor Analysis and Structural Equation Modeling, 2 ed. Cary, NC; SAS Institute.

Isakson K. and Jervis P., (1998): TheAdjustment of Adolescents during the Transition into High School: A Short-Term Longitudinal Study.*Journal of Youth and Adolescence,* Vol. 28.No. 1.

Jimmetra J. A., (2010): African American Males Perceptions of Factors Affecting Their Transition from Middle School to High School. Greensbore, 2010.Institute of Education, University of London 2008.

Research Division.(2008): A Study of *Students' Transition from Primary to Secondary Schooling*. New Zealand Ministry of education.

Timm, Neil H. (2002). *Applied Multivariate Analysis*. New York, NY: Springer

Warunga K. R., Musera G., and Sindabi O., (2011): Factors Affecting Transition Rates from Primary to Secondary Schools: The case of Kenya.*Problems of Education in the 21st Century*, Vol. 32. Pp. 129-139.

APPENDICES

Appendix 1

Correlation Matrix

		Literacy	De facto place of residence	Region	Religion	Ethnicity	Type of place of residence	Number of household members	Birth Index	Relationship to household head	Sex of household head	Has electricity	Main floor material	Main wall material	Main roof material	Has Water and Sanitation	Access to information
Correlation	Literacy	1.00	-.24	.094	-.30	.207	-.02	-.32	-.11	-.13	-.01	.130	.131	.095	.228	-.11	.189
	De facto place of residence	-.04	1.00	.301	.045	.261	.847	.157	.116	.029	.035	-.48	-.54	-.33	-.13	.511	-.23
	Region	.094	.301	1.00	-.28	-.04	.147	-.15	.230	-.07	-.03	-.21	-.10	-.34	.002	.141	.041
	Religion	-.30	.046	-.28	1.00	-.05	-.03	.130	-.02	.002	.124	-.09	-.27	-.03	-.14	.085	-.13
	Ethnicity	.007	.051	-.04	-.05	1.00	.051	.039	-.01	.02	-.03	-.08	-.19	-.11	-.10	.115	-.08
	Type of place of residence	-.02	.847	.147	-.03	.231	1.00	.177	.299	.063	.201	-.33	-.50	-.35	-.05	.497	-.17
	Number of household members	-.32	.157	-.15	.130	.039	.177	1.00	.221	.350	-.24	-.20	-.19	.200	-.14	.306	-.13
	Birth Index	-.11	.116	.230	-.02	-.01	.099	.221	1.00	.047	-.09	-.10	-.10	-.06	-.07	.194	-.22
	Relationship to household head	-.13	.026	-.07	.002	.00	.063	.350	.247	1.00	-.67	-.04	-.19	-.13	-.25	-.03	-.02
	Sex of household head	-.01	.035	-.03	.124	-.03	.001	-.24	-.09	-.67	1.00	.002	.127	.135	.201	.211	-.13
	Has electricity	.130	-.48	-.21	-.05	-.08	-.33	-.20	-.10	-.04	.002	1.00	.363	.181	.147	-.39	.334
	Main floor material	.131	-.54	-.10	-.07	-.09	-.60	-.19	-.10	-.15	.127	.363	1.00	.473	.288	-.43	.288
	Main wall material	.095	-.33	-.34	-.03	-.11	-.35	.020	-.06	-.13	.135	.181	.473	1.00	.647	-.36	.124
	Main roof material	.228	-.13	.200	-.14	-.11	-.09	-.14	-.07	-.25	.201	.147	.288	.547	1.00	-.21	.087
	Has Water and Sanitation	-.11	.511	.141	.085	.115	.497	.306	.194	-.02	.211	-.39	-.43	-.36	-.21	1.00	-.30
	Access to information	.189	-.23	.241	-.13	-.08	-.17	-.13	-.22	-.02	-.13	.334	.288	.194	.087	-.30	1.00
	Mode of transport	.100	.046	.070	-.01	-.06	.059	.055	-.03	.058	-.10	.083	-.26	-.15	.061	-.06	.227
	Preventive health measures	.021	.035	.113	.048	.186	-.05	.059	.155	.025	-.04	-.07	.056	.230	-.11	-.03	.161
	Type of cooking fuel	-.05	.686	.270	.022	-.08	.573	.154	.106	.035	.029	-.45	-.43	-.29	-.11	.390	-.24
	Wealth index factor score (5 decimals)	.185	-.71	-.14	-.05	-.09	-.73	-.34	-.19	-.16	.164	.519	.820	.505	.265	-.65	.435
	Parents marital status	.171	-.24	-.13	.008	.010	.002	-.26	-.05	-.11	.101	.185	.077	.270	.441	-.08	-.09
	Number of other wives	.100	-.11	-.19	.124	-.03	-.07	-.29	-.03	.185	-.07	.089	.057	.158	-.04	-.15	-.02
	Wife rank number	.100	-.11	-.19	.125	-.03	-.08	-.29	-.03	.185	-.07	.090	.059	.161	-.04	-.15	-.02
	Partner's education level	.301	-.24	-.05	.023	-.03	-.17	-.20	-.16	-.12	.081	.313	.290	.167	.163	-.29	.207
	Partner's Highest year of education	-.02	-.15	-.03	-.04	-.02	-.25	.039	.263	-.13	.135	-.04	.187	.136	.028	.209	-.11
	Partner's occupation	-.11	.046	.159	.002	-.02	-.03	-.15	-.06	.045	-.05	-.13	-.23	-.09	-.12	.032	-.12
	Mother's occupation	-.16	.045	-.05	-.02	-.14	.118	.028	-.05	-.07	.248	-.04	-.26	-.09	-.05	.261	-.01
	Mother Work for family, others, self	-.05	.172	-.03	-.08	.215	.102	.097	.292	.068	-.05	-.27	-.29	-.09	-.03	.196	-.05
	Mother Works at home or away	.029	-.12	-.03	-.02	-.05	.053	.052	.271	-.04	-.06	.168	-.21	.211	.113	-.15	-.05
	Who decided how to spend money	-.17	-.12	-.01	.060	-.08	-.02	.016	.204	.123	-.11	.090	-.23	-.03	-.07	-.17	.129
	Mother's type of earnings for work	-.05	.143	.121	-.10	-.01	.161	-.21	-.02	.005	-.07	-.01	-.12	-.06	-.04	-.05	.019

		Mode of transport	Preventive health measures	Type of cooking fuel	Wealth index factor score (5 decimals)	Parents marital status	Number of other wives	Wife rank number	Partner's education level	Partner's Highest year of education	Partner's occupation	Mother's occupation	Mother Work for family, others, self	Mother Works at home or away	Who decided how to spend money	Mother's type of earnings for work
Correlation	Literacy	.120	.021	-.05	.185	.171	.100	.100	.301	-.02	-.11	-.16	-.05	.029	-.17	-.05
	De facto place of residence	.046	.235	.685	-.71	-.04	-.11	-.11	-.24	-.15	.045	.045	.172	-.02	-.03	.143
	Region	.070	.113	.270	-.14	-.13	-.19	-.19	-.05	-.03	.159	-.21	-.03	-.03	-.01	.121
	Religion	-.21	.248	.022	-.09	.008	.124	.125	.023	-.04	.002	-.21	-.08	-.02	.160	-.10
	Ethnicity	-.16	.186	-.08	-.09	.010	-.13	-.03	-.09	-.03	-.02	-.14	.215	-.05	-.08	-.01
	Type of place of residence	.059	-.06	.573	-.73	.002	-.17	-.08	-.17	-.25	-.03	.118	.102	.053	-.02	.161
	Number of household members	.055	.069	.164	-.34	-.06	-.29	-.09	-.20	.239	-.05	.028	.297	.052	.206	-.01
	Birth Index	-.23	.155	.105	-.19	-.05	-.23	-.03	-.15	.263	-.05	-.21	.292	.071	.204	-.02
	Relationship to household head	.058	.226	.035	-.16	-.11	.139	.189	-.12	-.13	.043	-.27	.268	-.04	.113	.006
	Sex of household head	-.10	-.04	.025	.164	.101	-.17	-.07	.081	.135	-.05	.048	-.05	-.05	-.11	-.07
	Has electricity	.053	-.07	-.45	.519	.185	.059	.090	.318	-.04	-.13	-.24	-.27	.168	.150	-.01
	Main floor material	-.26	.056	-.43	.820	.077	.057	.069	.290	.187	-.03	-.26	-.09	-.01	-.03	-.12
	Main wall material	-.15	.230	-.29	.505	.270	.058	.061	.167	.136	-.09	-.25	-.09	.011	-.03	-.06
	Main roof material	.061	-.11	-.10	.265	.441	-.24	-.04	.163	.028	-.10	-.25	-.03	.113	-.07	-.04
	Has Water and Sanitation	-.26	-.03	.390	-.65	-.08	-.15	-.15	-.25	.209	.032	.051	.196	-.15	-.17	-.05
	Access to information	.227	.161	-.24	.435	-.09	-.23	-.02	.207	-.11	-.12	-.21	-.05	-.05	.139	.019
	Mode of transport	1.00	.063	.055	.218	.074	-.29	-.09	.013	-.05	-.21	-.13	.186	.035	.104	.054
	Preventive health measures	.053	1.00	.072	.204	-.03	-.21	-.21	.035	.226	-.09	-.16	.246	-.07	-.03	.075
	Type of cooking fuel	.056	.072	1.00	-.66	-.05	-.13	-.13	-.25	-.09	.089	-.21	.130	-.07	-.05	.128
	Wealth index factor score (5 decimals)	.018	.204	-.65	1.00	.081	.059	.062	.351	.138	-.09	-.26	-.17	.012	.016	-.06
	Parents marital status	.074	-.03	-.05	.081	1.00	-.28	-.08	.153	.117	-.03	-.21	-.07	.190	-.01	.028
	Number of other wives	-.29	-.21	-.12	.139	-.03	1.00	1.00	.001	-.03	.193	.053	-.12	.044	-.04	-.06
	Wife rank number	-.29	-.21	-.13	.163	-.08	1.00	1.00	.001	-.03	.193	.055	-.12	.044	-.04	-.06
	Partner's education level	.013	.035	-.25	.351	.153	.001	.001	1.00	-.19	-.14	-.27	-.11	.002	-.02	-.03
	Partner's Highest year of education	-.05	.226	-.09	.138	.117	-.03	-.03	-.19	1.00	.023	-.11	.204	-.07	-.01	-.04
	Partner's occupation	-.21	-.09	.089	-.09	-.03	.193	.193	-.14	.023	1.00	.370	-.06	-.01	-.13	.107
	Mother's occupation	-.15	-.16	-.01	-.26	-.07	.053	.062	-.07	-.11	.370	1.00	-.09	.089	-.08	.063
	Mother Work for family, others, self	.186	.246	.050	-.17	-.07	-.12	-.12	-.11	.204	-.06	-.25	1.00	-.27	-.15	.132
	Mother Works at home or away	.035	-.07	-.07	.012	.190	.044	.244	.002	-.07	-.01	.088	-.27	1.00	.029	-.10
	Who decided how to spend money	.124	-.02	-.05	.016	-.01	-.24	-.04	-.02	-.01	-.13	-.28	-.15	.029	1.00	.079
	Mother's type of earnings for work	.054	.275	.118	-.06	.028	-.26	-.06	-.03	-.04	.107	.053	.132	-.10	.079	1.00

Appendix 2

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared ...	
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance
1	5.488	17.703	17.703	5.488	17.703	17.703	4.930	15.904
2	2.607	8.410	26.113	2.607	8.410	26.113	2.142	6.910
3	2.174	7.014	33.127	2.174	7.014	33.127	2.050	6.612
4	1.899	6.127	39.254	1.899	6.127	39.254	2.049	6.609
5	1.734	5.594	44.848	1.734	5.594	44.848	1.736	5.600
6	1.477	4.766	49.614	1.477	4.766	49.614	1.664	5.369
7	1.416	4.567	54.181	1.416	4.567	54.181	1.584	5.108
8	1.285	4.147	58.328	1.285	4.147	58.328	1.487	4.797
9	1.183	3.818	62.145	1.183	3.818	62.145	1.297	4.184
10	1.097	3.540	65.685	1.097	3.540	65.685	1.258	4.057
11	1.062	3.427	69.112	1.062	3.427	69.112	1.228	3.962
12	.970	3.130	72.241					
13	.938	3.026	75.268					
14	.871	2.809	78.076					
15	.798	2.575	80.652					
16	.686	2.213	82.865					
17	.676	2.180	85.044					
18	.642	2.071	87.115					
19	.530	1.710	88.825					
20	.517	1.668	90.493					
21	.506	1.631	92.124					
22	.451	1.455	93.579					
23	.428	1.380	94.959					
24	.373	1.204	96.163					
25	.314	1.014	97.177					
26	.294	.949	98.126					
27	.240	.774	98.900					
28	.182	.588	99.488					
29	.099	.318	99.806					
30	.060	.194	100.000					
31	.000	.000	100.000					

Extraction Method: Principal Component Analysis.

Appendix 3

Rotated Component Matrix[a]

	Component										
	1	2	3	4	5	6	7	8	9	10	11
Wealth index factor score (5 decimals)	-.906										
De facto place of residence	.890										
Type of place of residence	.865										
Type of cooking fuel	.773										
Main floor material	-.745										
Had Water and Sanitation	.636										
Had electricity	-.556										
Wife rank number		.959									
Number of other wives		.958									
Main roof material			.815								
Main wall material	-.419		.729								
Parents marital status			.709								
Sex of household head				-.850							
Relationship to household head				.835							
Number of household members				.557	-.398						
Religion					-.772						
Literacy					.626						
Region					.601						
Partner's occupation						-.734					
Mother's occupation						-.644				-.401	
Mode of transport						.605					
Partner's Highest year of education							-.721				
Partner's education level							.600				
Access to information	-.387						.419				
Birth Index							-.394		-.380		
Mother Work for family, others, self								-.787			
Mother Works at home or away								.654			
Mother's type of earnings for work									.719		
Who decided how to spend money						.350		.386	.458		
Preventive health measures										.746	
Ethnicity											-.811

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 14 iterations.

Appendix 4

Regression Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	. 652[a]	.425	.399	.466

a. Predictors: (Constant), Ethnicity, Preventive Health Measures, Mothers Type Of Earnings, House Wife Status, Parents Education Level, Parents Occupation, Regional Influence, Family Composition, Home Environment, Family Position, Social Economic Status

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40.328	11	3.666	16.899	.000[b]
	Residual	54.669	252	.217		
	Total	94.996	263			

a. Dependent Variable: Highest educational level

b. Predictors: (Constant), Ethnicity, preventive health measures, Mother' s type of earnings , House wife status , Parent' s education level , Parent' s occupation , Regional influence , Family composition , Home environment , Family position , Social Economic status