Linear Discriminatory Function Approach to Literacy Rate Variation: Evidence from Africa

Jayadevan CM
Charles Stuart University, Australia
Email: jdevcm [AT] outlook.com

_____

ABSTRACT---- *Literacy plays an important role in everyone's daily life as it is an instrument for achieving higher education, employment and other accomplishments. Most social problems like exploitation, poverty, inequality, unemployment, prostitution, child labor and crime are the product of existence of illiteracy. The main purpose of this paper is to analyze the variation in average adult literacy rates among the countries of Africa on the basis of available statistical data for the period 1980-2013and shed some light for literacy rate management. Using the linear discriminant function analysis technique for countries in Africa for the period 1980-2013, the study has shown that significant discriminating factors responsible for the variation in literacy rate are secondary, tertiary and primary gender parity index (GPI) and the primary education starting age.*

Keywords---Literacy, adult, enrollment, primary, secondary, tertiary

_____

1. INTRODUCTION

Literacy plays an important role in everyone's daily life as it is the instrument for achieving higher education, employment and other accomplishments. Adult literacy in developing countries has lower rates than in developed countries, and the effects on social status are devastating. Illiterates are stuck in a rut of poverty and are incapable of working higher paying jobs. In developing countries where literacy rates are lower, poverty comes with challenges of education. Schools are few and far between and funds to operate the schools are just as scarce. Illiterate adults, especially women, struggle to find a voice in their communities, and severe gender discrimination affects millions.Without some education and critical thinking, people are highly susceptible to deception and manipulation to things like working a bad job with low pay, prostitution, child labor, and crime. Illiteracy holds people down in society and prevents communities from advanced thinking and technology. Everyday tasks are impossible and self-esteem reaches ultimate lows among illiterates. Literacy is the key to breaking out of poverty, putting a stop to prostitution and child labor, and reducing crime. It gives people the confidence to better their lives through education and pursue better jobs, and it gives them a voice and a place in today's society(Jennifer,2014).

The economic benefits of education to improve growth rates appear to be very large. A more educated society transalates into higher rates of economic growth and thus the ability of Governments to alleviate poverty(Arusha, 2009). Positive association between education quantity and economic growth has been observed in the studies of Hanushek(1995), Temple(2001), Behabib and Spiegel(1992). However, a weakassociation between education quantity and growth has been observed in the studies of Bilsand Klenow(200).

The main purpose of this paper is to analyze the variation in average adult literacy rates among the countries of Africa on the basis of available statistical data for the period 1980-2013. The average adult literacy rates have been classified into three categories. The study makes an attempt to find the factors responsible for the variation in average adult literacy rates. Higher the value of factors like enrollment ratio, trained teachers, persistence rate, education expenditure, higher the literacy rate. Similarly, higher the value of pupil-teacher ratio, repetition rate, education starting age, and education duration, lower the literacy rate. In other words, factors like enrollment ratio, trained teachers, persistence rate, education expenditure play a positive role in enhancing the literacy rate while factors like pupil-teacher ratio, repetition rate, education starting age and duration of education negatively impact the literacy rate.

2. MATERIALS AND METHODS

The main source of data are from Worldbank and UNESCO statistics.Following variables are used in our analysis. 1)TLR_15&above" (or adult Literacy rate), adult total (% of people ages 15 and above), 2)Prim.GPI:School enrollment, primary (gross), gender parity index (GPI)- Gender parity index for gross enrollment ratio in primary education is the ratio of girls to boys enrolled at primary level in public and private schools, 3)Sec.GPI:School enrollment, secondary (gross), gender parity index (GPI)- Gender parity index for gross enrollment ratio in secondary education is the ratio of

girls to boys enrolled at secondary level in public and private schools, 4)Ter.GPI:School enrollment, tertiary (gross), gender parity index (GPI)- Gender parity index for gross enrollment ratio in tertiary education is the ratio of women to men enrolled at tertiary level in public and private schools, 5)Prim.Starting Age:Primary school starting age (years), 6)Lower sec. school start. Age:Lower secondary school starting age (years), 7)Prim..Edu.,duration (years):Primary education, duration (years), 8)Sec.edu., duration:Secondary education, duration (years), 9)Pupil-teacher ratio, prim:Pupil-teacher ratio, primary, 10)Pupil-teacher ratio, sec: Pupil-teacher ratio, secondary, 11)Persis.rate, prim: Persistence to last grade of primary, total (% of cohort), 12)%Lower sec. compl. Rate:Lower secondary completion rate, 13)Repeaters, prim:Repeaters, primary, total (% of total enrollment), 14)%Total prim. trained teachers:%Total primary trained teachers15)Prim.exp as %edu exp: Current education expenditure, primary (% of total expenditure in primary public institutions), 16)Sec.exp as %edu exp: Current education expenditure, secondary (% of total expenditure in secondary public institutions), 17)Ter. exp as %edu exp: Current education expenditure, tertiary (% of total expenditure in tertiary public institutions), 18)Edu.exp.as %tot Govt exp: Government expenditure on education, total (% of government expenditure), 19)Govt. Edu.Exp.as %GDP: Government expenditure on education, total (% of GDP).

We made an attempt to find the significant factors responsible for the variation in literacy rates among countries of Africa using multiple regression but failed to show any significant results. In this paper we will use the linear discriminant analysis (LDA) as a technique for analyzing literacy rate variation. LDA is a statistical technique designed to investigate the differences between two or more groups of people with respect to several underlying variables. Because the variable being predicted is categorical, LDA technique is more appropriate than commonly used measures. LDA performs a multivariate test of differences between groups. In addition, LDA is used to determine the minimum number of dimensions needed to describe these differences.

LDA is used to analyze relationships between a dependent variable and independent variables. Adult literacy rate has been considered as the dependent variable. Since this is a continuous variable, this has been classified into three categories, that is 1)0-40, 2)above 40 to 70 and 3)above 70 to 100. LDA analysis attempts to use the predictor variables to distinguish among the groups of the response variable. If LDA is able to distinguish among groups, it must have a strong relationship to at least one of the predictor variables. Using LDA,a series of statistical tests are conducted to test the overall relationship among the predictor variables and groups defined by the response variable.

This paper is mainly concerned with an analysis to determine if there is a significant effect of factors like gross par, ity index, trained teachers, persistence rate, education expenditure, pupil-teacher ratio, repetition rate, education starting age and education duration on the literacy rate. There are 18 predictor variables.
The hypothesis of interest is:
$H_0: \beta_1 = \beta_2 = \beta_3 \dots = \beta_{18} = 0$; $H_a: Not\ all\ \beta_i\ equal\ zero$
This hypothesis has been tested using LDA. The test statistic used for LDA is $Wilk's\ Lambda\ \Lambda = \coprod_i \frac{1}{1+\lambda_i}$. Where λ_i are the eigen values of the corresponding design matrices. There are three main assumptions for LDA: they are 1)Multivariate Normality (MVN): To test for MVN, we begin by examining the marginal distributions of each univariate variable using box plots. If any of these plots show non-normality, then MVN is suspect and we use a procedure based on Mahalanobis distance, in which we construct a χ^2 probabilities to determine confirmity with multivariate normality. 2)Equality of covariances:the test for equality of covariances is based on Box's M-test and 3)Independence of observations: This test is a function of the experimental design, or data collection method and hence is not tested. For the purposes of this paper we assume that it is true.

3. EMPIRICAL RESULTS

The average adult literacy rate was 58.80% during the period 1980-2013. However, the adult literacy rates varied across contries of Africa. On the basis of average adult literacy rate, countries of Africa were divided into three categories, i.e. the adult literacy rate 1) 40 or below 40 2) Above 40, but upto 70 and 3) above 70. The average literacy rate was 81.41% for the third group, 56.72% for the second group and 28.51% for the first group (Table 1).

Countries like Uruguay, Eq.Guinea, Seychelles, Mauritius, Swaziland, Zimbabwe, Lesotho, Congo, Rep., Namibia, Botswana, Gabon, Libya, Kenya, Cabo Verde, Sao Tome&Principe, Tunisia and Cameroon had an average literacy rate above 70%. Countries like Angola, Tanzania , Sudan, Uganda ,Madagascar, Zambia, Egypt, Comoros, Congo, Dem. Rep.,Algeria, Eritrea, Burundi, Rwanda, Malawi, Togo, Nigeria, Morocco, Guinea-Bissau, Mauritania, Gambia, Senegal, Mozambique, Ghana, Central Africa, Sierra Leone, Liberia and Cote d'Ivoire had an average literacy rate between 40 to 70 during the same period. On the other hand, countries like Ethiopia, Benin, South Sudan, Mali, Chad, Burkina Faso and Niger had a literacy rate 40 or below. Country-wise average adult literacy rate is shown in Fig.1a.

Table 1: Summary Statistics of Adult Lietracy Rate

Statistic	Group 1	Group 2	Group 3	Total	Statistic		Group 1	Group 2	Group 3	Total
Mean	28.51	56.72	81.41	58.80	Median		28.70	57.09	79.94	61.26
Std. Deviation	7.72	8.83	8.46	22.01	Skewness	-.59	-.16	.56	-.22	
Minimum	10.89	40.98	70.20	10.89	Kurtosis		-.33	-1.13	-.70	-.89
Maximum	39.28	69.87	98.27	98.27						

Fig.1a:Average Adult Literacy Rates in Africa

Table 2:Correlations
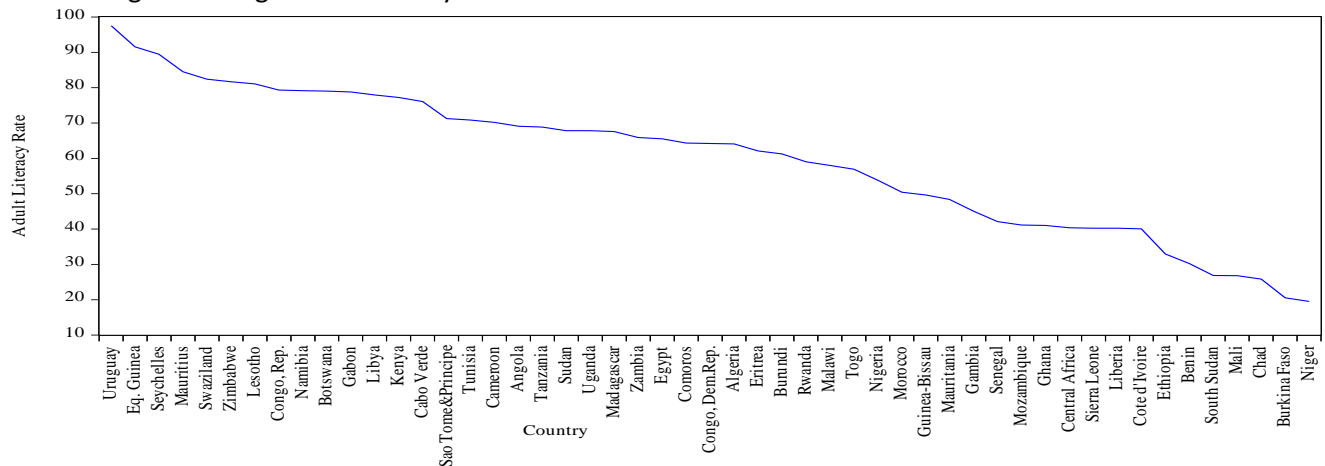
		TLR_15&above	Prim.GPI	Sec. GPI	Ter. GPI	Prim.Starting Age	Prim.Edu.duration	Pupil-teacher ratio, prim.	Persis.to last grade of prim.	Repeaters, prim.
TLR_15&above		1.00	.67**	.72**	.68**	-.39**	.18*	-.55**	.39**	-.22**

|	%Prim.trained teachers	Lower sec. school start.age	%Lower sec.compl.rate	Sec.edu., duration	Pupil-teacher ratio, sec.	Prim..exp.as %edu exp	Sec.exp.as %edu exp.	Ter.exp. as %edu exp.	Edu.exp.as %tot Govt.exp.	Govt. Edu. Exp.as %GDP
TLR_15&above	.21**	-.13	.62**	-.30**	-.41**	-.40**	.26**	.05	-.06	.297**

Significant positive correlation with adult literacy rate with secondary GPI, tertiary GPI and primary GPI shows that higher the enrollment ratio at secondary, tertiary and primary levels, higher the literacy rate. Significant negative correlation of adult literacy with pupil-teacher ratio at primary and secondary schools shows that higher the pupil-teacher ratio, lower the literacy rate. Significant positive correlarion of adult literacy rate with persistence rate and %lower secondary completion rate shows that adult literacy rate tend to be higher where persistence rate is high. Significant negative correlation of adult literacy rate with primary starting age shows that literacy rate tend to be high where primary starting age is low. Similarly, the positive correlation of Government expenditure as % of GDP with literacy rates shows that higher the Government expenditure as % of GDP, higher the literacy (Table 2).

Primary, secondary and tertiary gross parity index for group 3 and group 2 are higher than group 1. Primary starting age is lower for groups 3 and 2. Pupil-teacher ratio at primary and secondary levels are far lower for group 3. Primary persistence rate is higher for group 3 than groups 1 and 2. % secondary completion rate is higher for group 3 than groups 1 and 2. Primary repeating rate is higher for group 1 than groups 2 and 3. %of primary trained teachers is higher for group 3 than other groups. Government educational expenditure as % of GDP is higher for group 3 than other groups(Table 3).

Table 3:Group Statistics

|	Group 1			Group 2			Group 3			Total		
	Mean	Std. Dev.	Valid N	Mean	Std. Dev	Valid N	Mean	Std. Dev	Valid N	Mean	Std. Dev	Valid N
Primary GPI	0.72	0.11	33	0.89	0.13	49	0.95	0.10	33	0.86	0.15	115
Secondary GPI	0.56	0.14	33	0.74	0.17	49	1.00	0.19	33	0.76	0.24	115
Tertiary GPI	0.35	0.23	33	0.50	0.23	49	1.02	0.39	33	0.61	0.39	115
Prim. Starting Age	6.64	0.49	33	6.39	0.49	49	6.03	0.47	33	6.36	0.53	115
Lower sec. school start. Age	12.61	0.50	33	12.55	0.79	49	12.36	.82	33	12.51	0.73	115
Prim..Edu., duration	5.97	0.17	33	6.16	0.59	49	6.33	0.54	33	6.16	0.51	115
Sec.edu., duration	6.73	0.45	33	6.29	0.71	49	6.09	0.88	33	6.36	0.74	115
Pupil-teacher ratio, prim.	50.58	12.76	33	43.77	13.32	49	31.26	11.46	33	42.13	14.60	115
Pupil-teacher ratio, sec.	28.70	8.58	33	25.63	10.18	49	19.57	5.61	33	24.77	9.26	115
Persis.to last grade of prim	57.03	15.99	33	58.79	17.26	49	73.69	22.78	33	62.56	19.84	115
%Lower sec. compl. Rate	16.17	10.93	33	28.50	17.00	49	53.17	25.42	33	32.04	23.33	115
Repeaters, prim.	17.89	7.14	33	14.86	9.05	49	11.38	7.40	33	14.73	8.38	115
%Total prim. trained teachers	73.43	16.14	33	76.60	20.99	49	83.48	16.24	33	77.66	18.66	115
Prim.exp as %edu exp	49.13	12.34	33	43.38	9.73	49	38.87	10.96	33	43.74	11.47	115
Sec.exp as %edu exp	26.83	11.49	33	30.56	10.11	49	34.88	9.93	33	30.73	10.82	115
Ter. exp as %edu exp	18.93	6.06	33	18.66	7.22	49	17.82	7.93	33	18.50	7.08	115
Edu. exp. as %tot Govt exp	15.38	4.17	33	16.40	5.31	49	18.10	7.10	33	16.59	5.66	115
Govt. Edu.Exp. as %GDP	3.32	1.25	33	4.08	1.36	49	5.81	2.73	33	4.36	2.07	115

Group-wise box plots for different variables are shown below. For TLR_15&above, extreme values are not observed, but higher median is observed for groups 3 and 2. However, higher variance is observed for groups 3 and 1 (Fig.1). For primary GPI, higher median, variance and extreme values is observed for groups 3 and 2 (Fig.2) . For secondary GPI, higher median and variance is observed for groups 3 and 2, but extreme values is observed for groups 1 and 3 (Fig.3).

Fig.1 TLR_15&above

Fig.2 Primary GPI

Fig.3 Secondary GPI
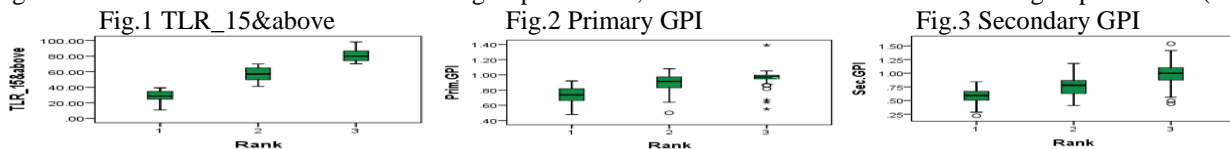
For tertiary GPI, higher median and variance is observed for groups 3 and 2, but extreme value is observed for group 1 (Fig.4). For primary pupil-teacher ratio, higher median is observed for group 1, but higher variance is observed for groups 1 and 2, but no extreme value is observed for any group(Fig.5). For primary repeating rate, higher median is observed for group 1, but higher variance is observed for groups 2 and 3, but extreme value is observed for group 3(Fig.6).

Fig 4 Tertiary GPI

Fig 5 Pupil-Teacher Ratio,prim.
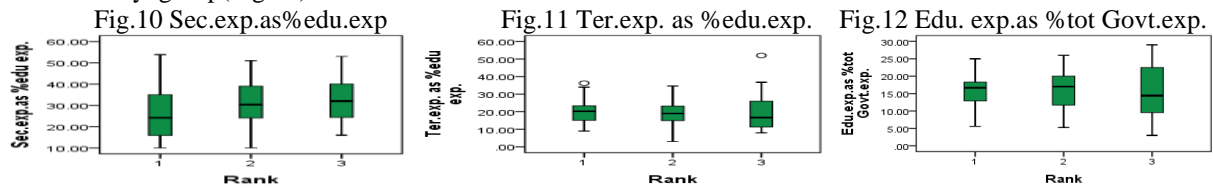

Fig 6 Repeating rate, prim.
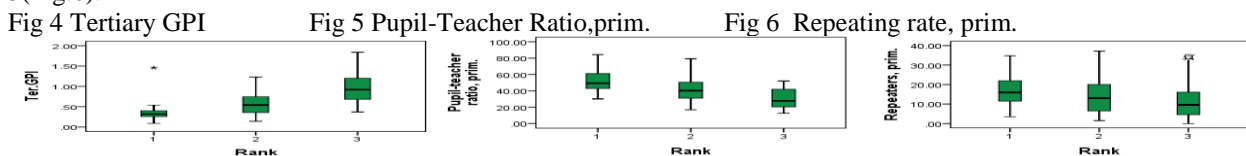
For primary trained teachers, higher median is observed for groups 3 and 2, but higher variance is observed for groups 2 and 3, but no extreme value is observed for any group(Fig.7).For secondary pupil-teacher ratio, higher median and variance is observed for groups 1 and 2, but extreme values are observed for all groups(Fig.8). For primary expenditure as % of education expenditure, higher median is observed for groups 1 and 2, but no extreme values are observed for any group(Fig.9).

Fig.7 %Prim.Trained Teachers

Fig.8 Pupil-Teacher Ratio, Sec

Fig.9 Prim.exp.as %of edu.exp
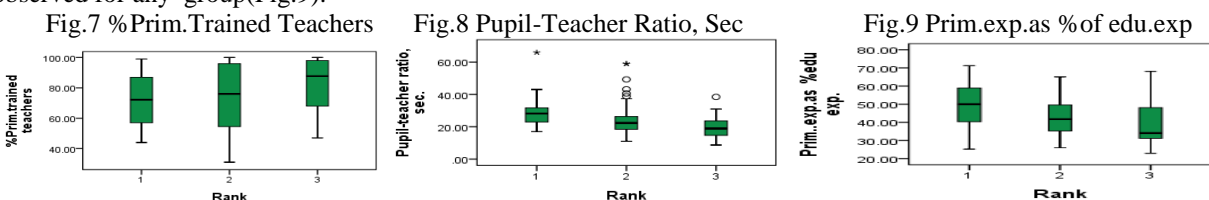
For secondary expenditure as % of education expenditure, higher median is observed for groups 2 and 3, but higher variance is found for group 1 and no extreme values are observed for any group(Fig.10). For tertiary expenditure as %

of education expenditure, higher median is observed for group 1, but higher variance is found for group 2 and extreme values are observed for groups 1 and 3(Fig.11). For education expenditure as % of total Government expenditure, higher median is observed for group 1, but higher variance is found for groups 3 and 2 and no extreme values are observed for any group(Fig.12).

Fig.10 Sec.exp.as%edu.exp   Fig.11 Ter.exp. as %edu.exp.   Fig.12 Edu. exp.as %tot Govt.exp.

For Government education expenditure as % of GDP, higher median is observed for groups 3 and 2, but higher variance is found for group 3 and 1, however, extreme value is observed for group 3(Fig.13).

Fig.13 Govt.Edu.Exp as % of GDP
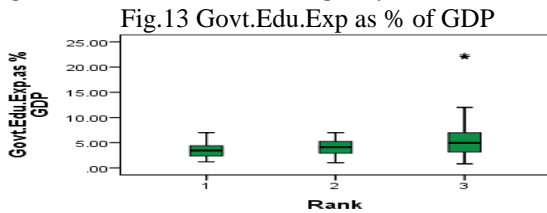
Table 4:Analysis Case Processing Summary

Unweighted Cases		N	Percent
The Valid		115	65.3
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	61	34.7
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	61	34.7
Total		176	100.0

The minimum ratio of valid cases to independent variables for LDA is 5 to 1. In this case, it is $115/18 \approx 6$ to 1, which satisfies the minimum requirement. However, it does not satisfy the preferred ratio of 20 to 1(Table 4).

Table 5:Prior Probabilities for Groups

Rank	Prior	Cases Used in Analysis	
		Unweighted	Weighted
1	.287	33	33
2	.426	49	49
3	.287	33	33
Total	1.000	115	115

The number of cases in the smallest group in this problem is 33, which is larger than the number of predictor variables (18), satisfying the minimum requirement. In addition, the number of cases in the smallest group satisfies the preferred minimum of 20 cases (Table 5).

Table 6:Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.779[a]	86.8	86.8	.800
2	.271[a]	13.2	100.0	.462

In this analysis there were 3 groups defined by category of literacy rates, 18 independent variables, so the maximum possible number of discriminant functions was 2. The canonical correlations for the dimensions one and two are 0.800 and 0.462, respectively (Table 6).

Table 7:Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.283	139.468	8	.000
2	.787	26.518	3	.000

In the table of Wilk's lambda whih tested functions for statistical significance, the stepwise analysis identified 2 discrimiinant functions that were statistically significant. The Wilk's lambda statistic for the test of function 1 through 2 functions (chi-square=139.47) had a probability of 0.000 which was less than the level of significance of 0.05. The Wilk's lambda statistic for the test of function 2 (chi-square=26.52) had a probability of 0.000 which was less than the level of significance of 0.05. The significance of the maximum possible number of discriminantfunctions supports the interpretation of a solution using 2 discriminant functions(Table 7).

Table 8:Functions at Group Centroids

Rank	Function	
	1	2
1	-1.590	-.521
2	-.182	.592
3	1.860	-.359

Table 8 shows unstandardized canonical discriminant functions evaluated at group means. Function 1 separates the literacy rate category 3(the positive value of 1.860) from literacy rate category 1(negative value of -1.590) and literacy category 2(negative value of -0.182). Function 2 separates the literacy rate category 2(the positive value of 0.592) from literacy rate category 1(negative value of -0.521) and literacy category 3 (negative value of -0.359).

Table 9:Variables Entered/Removed[a,b,c,d]

Step	Entered	Min. D Squared					
		Statistic	Between Groups	Exact F			
				Statistic	df1	df2	Sig.
1	Sec.GPI	1.121	1 and 2	22.102	1	112.000	7.417E-6
2	Prim.GPI	2.290	1 and 2	22.380	2	111.000	6.827E-9
3	Prim.Starting Age	3.136	1 and 2	20.242	3	110.000	1.614E-10
4	Ter.GPI	3.220	1 and 2	15.451	4	109.000	4.844E-10

When we use the stepwise method of variable inclusion,we limit our interpretation of predictor variables to those listed as statistically significant in the table of variables Entered/Removed. We will interpret the impact on membership in groups defined by the response variable by the predictor variables:1)Secondory GPI 2)Primary GPI, 3)Tertiary GPI and 4)Primary starting age (Table 9).Differences in literacy rate observed between groups 1 and 2 is mainly caused by the higher gross parity index at secondary, tertiary and primary levels of education as well as lower primary education starting age.

Using Wilk's lambda and step-wise LDA, the variables that minimizes the overall Wilk's lambda is entered. In our case, secondary GPI, Tertiary GPI, Primary GPI, primary starting age, primary teacher-pupil ratio, %lower sec completion rate and Government Education Expenditure as %GDP are significant (Table 10).

Table 10:Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Primary GPI	0.61	36.50	2.00	112.00	0.000
Secondary GPI	0.50	57.01	2.00	112.00	0.000
Tertiary GPI	0.52	51.33	2.00	112.00	0.000
Primary Starting Age	0.81	13.11	2.00	112.00	0.000
Primary.Education, duration (years)	0.93	4.53	2.00	112.00	0.013
Pupil-teacher ratio, primary.	0.74	19.94	2.00	112.00	0.000
Persis.to last grade of prim., tot.(% of cohort)	0.87	8.30	2.00	112.00	0.000
Repeaters, prim., total (% of tot. enr.)	0.91	5.36	2.00	112.00	0.006
%Total prim. trained teachers	0.96	2.60	2.00	112.00	0.079
Lower sec. school starting age (years)	0.98	1.03	2.00	112.00	0.362
%Lower sec. completion rate total	0.62	34.54	2.00	112.00	0.000
Sec.education, duration (years)	0.89	7.20	2.00	112.00	0.001
Pupil-teacher ratio, secondary	0.85	9.68	2.00	112.00	0.000
Primary.exp as %edu exp	0.88	7.38	2.00	112.00	0.001
Secondary.exp as %edu exp	0.92	4.88	2.00	112.00	0.009
Tertiary. exp as %edu exp	1.00	.22	2.00	112.00	0.801
Education expenditure as %tot gov exp	0.97	1.99	2.00	112.00	0.141
Government Education Expenditure as %GDP	0.78	16.20	2.00	112.00	0.000

Table 11:Structure Matrix

Variables	Function 1	Function 2	Variables	Function 1	Function 2
Sec.GPI	.756*	-.006	Edu.exp.as %tot Govt.exp.[b]	.168*	.015
Ter.GPI	.700*	-.408	%Prim.trained teachers[b]	.146*	-.019
%Lower sec.compl.rate[b]	.427*	.091	Prim.GPI	.547	.663*
Prim.Starting Age	-.363*	-.003	Prim.Edu.duration[b]	.097	.320*
Govt.Edu.Exp.as %GDP[b]	.340*	-.068	Repeaters, prim.[b]	-.200	-.294*
Pupil-teacher ratio, prim.[b]	-.320*	.135	Sec.exp.as %edu exp.[b]	.132	-.250*
Pupil-teacher ratio, sec.[b]	-.234*	.034	Lower sec. school start.age[b]	-.175	.214*
Persis.to last grade of prim.[b]	.176*	-.050	Prim..exp.as %edu exp.[b]	-.138	.188*
Sec.edu., duration[b]	-.168*	.050	Ter.exp. as %edu exp.[b]	.062	.157*

Based on the structure matrix, the predictor variables strongly associated positively with discriminant function 1 which distinguished between literacy rate categories are Secondary GPI(r=0.756) and Tertiary GPI(r=0.700) and negatively associated with primary starting age(-0.363).Based on the structure matrix, the predictor variable strongly associated positively with discriminant function 2 which distinguished between literacy rate categories is Primary GPI(r=0.663).Other predictor variables strongly associated with discriminant function 1 which were strongly associated with literacy rates were %Lower sec. completion rate(r=0.427), Government Education Expenditure as %GDP (0.34) and Pupil-teacher ratio, primary(r=-0.32)(Table 11).

**Table 12:Standardized Canonical Discriminant Function Coefficients**

	Function 1	Function 2
Prim.GPI	.165	1.076
Sec.GPI	.656	-.178
Ter.GPI	.264	-.700
Prim.Starting Age	-.630	-.097

The number of discriminant dimensions is the number of groups minus 1. However, some discriminant dimensions may not be statistically significant. In this example, there are two discriminant dimensions, both of which are statistically significant. The Coefficients of linear discriminants are reported in Table 12. The equations of the linear discriminante function are:

1)discriminant_score_1=0.656*Sec.GPI+0.264*Ter.GPI+0.165*Prim.GPI-0.63*Prim.Starting.Age
2)discriminant_score_2 =1.076*Prim.GPI – 0.178 Sec.GPI - 0.700*Ter.GPI -0.097 Prim.starting Age).

As you can see, the literacy rate category 1 tend to be at the more primary starting age (negative) end of dimension 1. The literacy rate category 3 tend to be at the opposite end in the dimension one and the literacy category 2 in the middle. On dimension 2, the literacy rate category 2 tend to be higher on Primary GPI and literacy category 1 and 3 lower(Fig 14).

Fig.14:Canonical Discrimination Functions
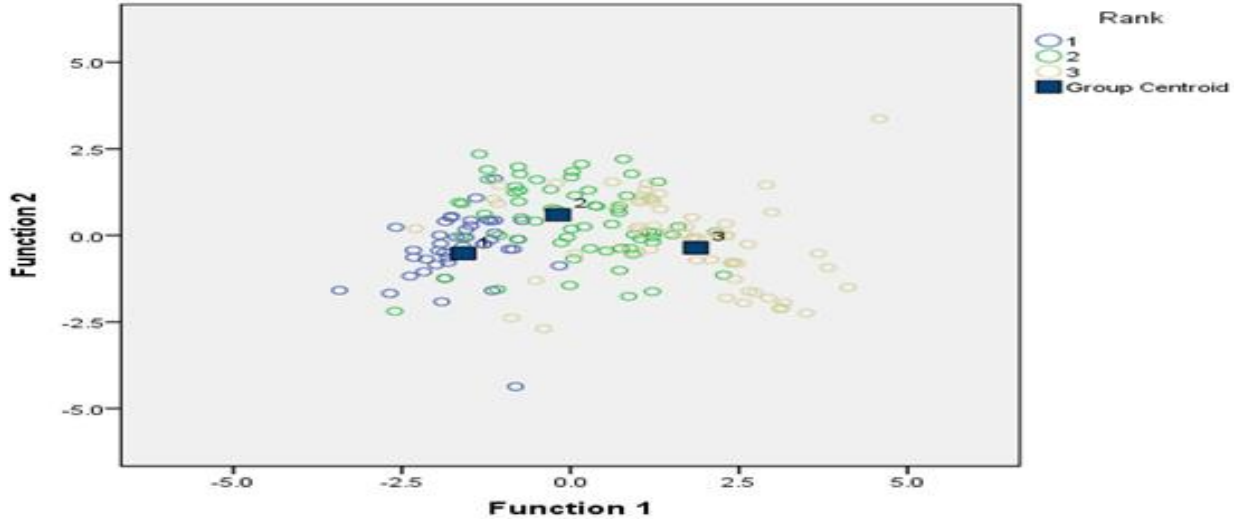
The cross validated accuracy rate computed by SPSS was 67.8% which was greater than or equal to the proportional by chance accuracy criteria of 43.7% (1.25*35.0=43.7). The criteria for classification accuracy is satisfied(Table 13). The proportional by chance accuracy rate was computed by squaring and summing the proportion of cases in each group from the table of prior probabilies for groups (0.287^2 + 0.426^2 + 0.287^2 =35.0).

Table 13:Classification Results[a,c]

|		Rank	Predicted Group Membership 1	2	3	Total
Original	Count	1	32	10	0	42
		2	6	45	14	65
		3	4	18	42	64
	%	1	76.2	23.8	.0	100.0
		2	9.2	69.2	21.5	100.0
		3	6.3	28.1	65.6	100.0
Cross-validated[b]	Count	1	31	11	0	42
		2	6	45	14	65
		3	4	20	40	64
	%	1	73.8	26.2	.0	100.0
		2	9.2	69.2	21.5	100.0
		3	6.3	31.3	62.5	100.0

Note:a. 69.6% of original grouped cases correctly classified.,
 b. Cross validation is done only for those cases in the analysis. In cross validation, each
 case is classified by the functions derived from all cases other than that case.
 c. 67.8% of cross-validated grouped cases correctly classified.

Apart from linearity the main assumptions in lda are:

1)MVN errors: The first assumption can be checked using Mahalanobis plot although symmetry is probably more important. If normality can not be induced by transformation or if the data are seriously non normal ie categorical, then the alternative of logistic regression should be used. It is worth pointing out that if all the assumptions are satisfied, lda is the optimal procedure and so should be used.

Fig.15:Normal Q-Q Plot for Multivariate DataFig.15a:Box Plot for p_mh
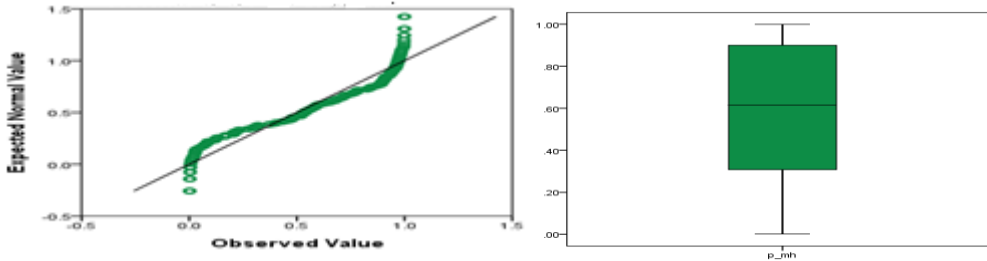
The plot of ordered Mahalanobis distances against their expected values under the assumption of Multivariate Normality clearly shows slight deviation from the straight line at the bottom and top portion. However, there is normality in the middle region. So we conclude that the assumption of multivariate normality is approximately upheld.The distribution is negatively skewed(Fig.15a).

2) **Box's Test of Equality of Covariance Matrices**.
For the second assumption there is a test of equality of covariances matrices, Box's M test. Violation of this assumption can affect significance tests of classification results. The significance level can be inflated (false positives) when the number of variables is large and the sample sizes of the groups differ. Quadratic methods can be used if the covariance matrices are unequal but a large number of parameters are involved and lda is thus superior for small sample sizes. Overall lda is robust to both the assumption of MVN and equality of covariance matrices, especially if the sample sizes are equal. The formal hypothesis forBox's M test for Equality of covariancewould be:$H_0: \sum 1 = \sum 2 = \sum 3$, $H_0: \sum 1 \neq \sum 2 \neq \sum 3$

$\alpha = 0.05$, $F_{obs} = \frac{MS_{Regression}}{MS_{Residual}}$

Reject H_0 if p-value <0.05
Do not reject H_0 as p-value $= 0.103 > 0.05$

Table 14:**Test Results**

Box's M		29.911
F	Approx.	1.415
	df1	20
	df2	35612.320
	Sig.	.103

Test Statistic
$$M = \sum n_i \ln|s| - \sum_{i=1}^{k} n_i \ln|s_i|$$
$$C^{-1} = 1 - \frac{2p^2+3p-1}{6(p+1)(k-1)} (\sum_{n-1}^{k} \frac{1}{n_i} - \frac{1}{\sum n_i})$$
Sampling Distribution
$$MC^{-1} \sim \frac{\chi^2 (k-1)(p)(p+1)}{2} \; if \; k, p < 5 \; and \; n_i \approx 20 \; else \; F \; distribution$$

To test the assumption of Equality of Co-variances, we use Box's M-test. If the Box's M Test shows p <.05, the covariances are significantly different and the null hypothesis is NOT rejected. If the Box's M Test shows p >.05, the covariances are not significantly different and the null hypothesis is not rejected. The value of Box's M is 29.91, with a p-value of 0.103, indicating that the assumption of equal co-variances is satisfied and null hypothesis is not rejected. So the assumption of homoscedasticity is not violated. . That is we do not reject the null hypothesis of $H_0: \sum 1 = 2=3$.Thus, both assumptions, namely, multivariate normality and equality of covariance matrices are satisfied.

4. CONCLUSION

Using the LDA technique for countries in Africa for the period 1980-2013, the study has shown that significant discriminatory factors responsible for the variation in literacy rate are Secondary GPI, tertiary GPI, primary GPI and Primary education starting age.In other words, most discriminating factors of adult literacy rate are gender parity index for gross enrollment ratio at secondary, tertiary and primary education levels. There is very high disparity in the ratio of girls to boys enrolled at secondary,tertiary and primary level in public and private schoolsacross countries. Similarly, the education starting age at primary level in countries where literacy rate is low is comparatively high. So in order to achieve higher literacy rate for countries in first and second rank categories, gender parity index for gross enrollment

ratio at secondary, tertiary and primary education levels need to be increased. Also, primary education starting age need to be reduced from 8 or 7 years to 5 years.

5. REFERENCES

[1] Arusha, V Cooray, "The Role of Eductaion in Economic Growth", Research Online,University of Woollongong, pp1-3,2009.

[2] Benhabib, J and Spiegel M, "The Role of Human Capital in Economic Develpoment: Evidence from Aggregate Cross Country Data.",Journal of Monetary Economics, 34, pp.143-173, 1994.

[3] Bils M and Klenow P, "Does Schooling Cause Growth?" American Economic Review, 90, pp.1160-1183,2000.

[4] Hanushek E, "Interpreting Recent Research on Schooling in Developing Countries", World Bank Research Observer, 10, pp.227-246.1995.

[5] Jennifer Flynt, http://blog.alfalit.org/the-importance-of-adult-literacy-in-todays-society, pp.1,2014.

[6] Temple J "Growth Effects of Education and Social Capital in OECD Countries", Economic Studies, 33, pp.57-101,2001.

[7] Worldbank, WorldBank Statistics,2015.