REGRESSION ANALYSIS OF ENTRY BEHAVIOUR

AND FINAL PERFORMANCE AND ONE WAY

ANOVA OF KENYAN SCHOOLS:

A Case Study of Nyamira District, Nyamira County

By

ONDIMA, Cleophas Mecha

A Project report Submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Applied Statistics

School of Mathematics, Statistics and Actuarial Science

MASENO UNIVERSITY

©2014

ABSTRACT

The performance in Kenyan Certificate of Primary education (KCPE) is very crucial in determining the learner's final grade in Kenya Certificate of Secondary the learners get admitted to, they can improve, maintain or drop. It is known that the individual's life structure is shaped by three types of external events namely, the social cultural environment, the role they play and the relationships they have and opportunities and constraints that enable them express and develop their personality. The main purpose of the study was to analyze entry behaviour and final performance and one way Anova to find out whether the true relationship existed. The specific objectives of the study were to analyze the relationship between entry marks and final grade, to verify whether entry marks impacted on the final performance, to test the null hypothesis that the entry marks had a direct impact on KCSE and to test the null hypothesis that the 15 schools selected had differences in their means at 1% and 5% significance levels. The study was carried out in Nyamira district public secondary schools in Nyamira County which are National, provincial and district. The data used in this study comprised of 48 schools of which a sample of 15 schools was used to select a sample of the accessible schools. A sample of 572 students of which 348 were girls and 224 were boys was analyzed. Given the reliability of regression analysis and one way Anova, the result would be used to alert the management, stakeholders and parents the level of learners and the strategies needs to achieve desirable results. Furthermore, it will let educators know the existing situation and suggest away forward to improve the situation.

CHAPTER 1

1.1 Background of the Study

The general problem of finding equations of approximating curves which fit given sets of data is known as curve fitting. The type of equation is often suggested from the scatter diagram. We could use a straight line $y = a + bx$

Many applications of regression involve situations where we have more than one regression variable taking the following form,

$$N_p = M_0 + M_1 x_1 + M_2 x_2 + - - - + M_q x_q + \varepsilon$$

with N_p being predicted value, x_i's, the predictor and M_i's the coefficients of predictors. ε is the error term which is the independent variable of the model but not considered. The dependent variable is denoted by N_p. The predictors considered are:

x_1 = entry behavior (eb)

x_2 = end of form three exams (ef3e)

x_3 = Mock exams (me)

x_4 = gender (g)

The regression models are often used as approximating functions i.e. the relationship between N_p and x_1, x_2, - - -x_q is not predictable but over certain independent variables the linear regression model is an approximation of the above variables. We can use multiple regression techniques to analyze the models as follows.

$$N_p = M_0 + M_1x_1 + M_2x_2 + M_3x_3 + M_4x_4 + \varepsilon$$

The focus on this study was to examine how the entry behavior affects the final performance in Nyamira district, Nyamira County by finding statistically the effect of entry behavior, end of form three examinations, mock examinations and gender.

There are many extrinsic factors such as availability of resources, teachers' qualifications and teaching experience, the learners' socio-economic background and teachers' motivation which contribute very much to learners' performance.

Students' attitude towards a subject greatly has a bearing in influencing the performance. It affects the individual's organized manner of thinking, feelings and reacting to study subject [8]. An individual's attitude towards a subject in African schools will influence their self concept of academic ability [9]. The significant other namely the teachers and peers have a great impact in the development of a student's attitude towards a subject [16]. The bulk of the studies have been on Mathematics and Sciences among primary and secondary school students. These subjects are regarded as hard by many students and are the cause of poor overall performance.

The European schools regard teacher's performance as that influenced by the management. Influence of teachers in an

institution has been an issue since early days of organizational theory with [21] prescribing highly organized structures and the most efficient use of resources. European and worldwide operations [11] are searching for more internationally coordinated standardized and justified approach to influencing *teacher to maintain better standards and improve the performance* of the learners.

The career of learners is based on the grade that is achieved at the *end of KCSE. This career is established by imparting on learner's* initiatives, innovations and skills as they start in form one. The firm foundation put in form o ne completely reverses the trend of performance in KCSE. Based on what had been analyzed, there was need to explore ways through which this was to bę do ne by analyzing entry behavior and final performance and one way Anova in secondary schools in Nyamira district, Nyamira County.

1.2 Statement of the Problem

Students who score good marks at KCPE and up scoring good grades at KCSE, those who score lowly can work hard and pass and some maintain. This state posses a great concern to all stakeholders because of unpredictable end results of the learners. Curriculum innovations at secondary level have not fully addressed this issue as evidenced by the continued unpredictable. There have been many reports following unpredictable at primary

and secondary levels and little has been done to know the causes of this poor performance. There is always an assumption that once a student has been admitted to secondary, he/she can perform regardless of the marks scored. This is assumed that through interaction with course content and completion of the syllabus, there are always better results. So long as the curriculum continues to emphasize the teaching of the core subjects for prospective students, it is imperative to ensure that the products of the process are well qualified to pursue their career path. Regression analysis of entry behavior and final performance in secondary school in Nyamira district and a one way Anova will enable us establish the cause of unpredictable performance in KCSE with reference to KCPE. This will be done by using the information received from KCPE of 572 students of which 348 are girls and 224 are boys.

1.3 Objectives of the Study

The general objective is to analyze entry scores and final performance using regression and one way Anova. The specific objectives will be:-

i) To analyze the relationship between entry marks and final performance.

ii) To verify whether entry marks impacts on the final performance.

iii) To determine the extent to which entry marks can influence the final performance.

1.4 Assumptions

In this study the following assumptions were made:

a) Students in the schools under study had similar entry behaviour.

b) The schools selected for the study were of similar socio-economic background

c) Syllabus coverage in all the schools under study was satisfactory

d) Use of marks and grades will render the results accurate.

e) The findings of the study, even though they were restricted to Nyamira district, Nyamira County, applies to other Counties in the country and the rest of the world.

1.5 Significance of the Study

To achieve better academic results, research findings are relied upon during education decision making processes. This study was expected to contribute to the generation of helpful information that was supposed to be accurate to education decision makers by its usefulness of recognizing the nature of this analysis. Given that entry marks and final performance were taken into consideration in modeling process, there was a change of approach in establishing determinants on the level of

performance. Taking into consideration the grade learners achieved at KCPE level and other factors, educators can put policies in place to make the whole initiative achieve the desired results. This influenced policy makers to tackle education in a way that will enable them to seek viable interventions for better manpower development and performance.

1.6 Basic Concepts

a) Regression

It is a type of analysis used when a researcher is interested in finding out whether an independent variable predicts a given dependent variable [15]. Regression is categorized into:-

i)   Univariate (simple regression)

ii)   Multivariate (multiple regression)

i) Univariate (simple) regression

It is used when the researcher is dealing with only one dependent variable and one independent variable [14]. The researcher might be interested in finding out whether entry scores predict the final grade. Final grade will be the dependent variable and entry scores will be the independent variable. That is

$y = a + bx$

where y represents the final performance 'a' represents regression constant and 'x' represents entry scores and 'b' represents constant for entry scores

ii) Multivariate (multiple) regression)

Multivariate (multiple) regression attempts to determine whether a group of variables together predict a given dependent variable the researcher was interested to find whether entry scores, mocks and end of form three exams influence the final grade in form four. The four independent variables are considered altogether in one equation. That is

$$Y = a + b_1x_1 + b_2x_2 - - - b_nx_n.$$

This reduces to:

$$Y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \varepsilon$$

Regression refers to the statistical methodology for predicting values of one or more (dependant) variables, N_{pi}'s from a collection of independent variables x_i's.

The equation is defined that for a given value of variable $X = x$, the actual value is determined by the expression $M_o + M_1x_1$ and some random variable caused by unmeasurable factors. If M_o is known as the true intercept and M_1 the true slope then N_p can be predicted within some random error ε. The relationship between x and N_p is represented by a straight line incase of a linear regression.

The regression model states that N_p is composed of mean, which depends in a continuous manner on the x_i's and the random variable ε which is the measurement error and other variables not considered in the model. The bivariate regression (simple regression) N_p is a function of only one independent variable i.e

$N_p = M_o + M_1 x + \varepsilon$

For a given value of the variable $X = x$, the value of N_p is determined by the expression $M_o + M_1 x$, and random variable ε caused by unmeasured factors. If we know the value M_o, the true population slope we can predict the value of N_p within some random error ε.

In multiple regression models, N_p is a function of two or more independent variables i.e

$N_p = M_o + M_1 x_1 + M_2 x_2 + \; - \; - \; - \; + M_q x_q + \varepsilon$ which is caused by the unmeasured factors.

The coefficients $M_1, M_2 \; - \; - \; -M_k$ are similar to the slope coefficient N_1 in the univariate model,

$N_p = M_o + M_1 x_1 + \varepsilon$

Each of the M_i coefficients represents the change in the dependent variable, N_p, if the variable associated with the coefficient of interest x_i, is changed by one unit and all other variables in the model are held constant.

Multiple regression equation describes how the mean value of N_p is related to x_1, x_2, x_k.

Since $E(\varepsilon) = 0$ hence

$E(N_p) = M_o + M_1 X_1 + M_2 X_2 + \; - \; - \; - \; + M_q X_q$

One way analysis of variance:

Sampling theory is used to test the significance of differences between two sampling means. In some situations there is need to

test the significance of difference among three of more sampling means or to test the null hypothesis that the sample means are all equal. A one way factor experiment measurements are obtained for an independent groups of samples where the number of measurements in each group b. We speak of **a** treatments each of which has **b** repetitions or replications. The results of **a** one-factor experiment can be presented in a table having **a** rows and **b** columns (**Table 1.10**).

x_{jk} denotes the measurement in j^{th} row and k^{th} column, where $j=1, 2, ---, a$ and $k = 1,2, ---, b$.

Table 1.10 (one factor experiment table)

Treatment 1	$x_{11}\ x_{12}\ ---\ x_{1b}$	$\bar{x}_1$
Treatment 2	$x_{21}x_{22}\ ---\ x_{2b}$	$\bar{x}_2$
Treatment a	$x_{a1}\ x_{a2}\ ---\ x_{ab}$	$\bar{x}_a$

We denote by $\bar{x}_{j.}$ the mean of the measurements in the jth row. We have

$$\bar{x}_{j\cdot} = \frac{1}{b}\sum_{k=1}^{b} x_{jk} \qquad j=1, 2, ---, a \tag{1}$$

The dot in $\bar{x}_{j\cdot}$ is used to show that the index k has been summed up. The values $\overset{a}{x}\overset{b}{j}_{\cdot}$ are the group means, treatment means or row means. The grand mean or overall mean is the mean of all measurements in all the groups and is denoted by $\bar{x}$, i.e.

$$\bar{x} = \frac{1}{ab}\sum_{j,k} x_{jk} = \frac{1}{ab}\sum_{j=1}\sum_{k=1} x_{jk} \tag{2}$$

Total variation, variation within treatments,

Variation between treatments

Total variation, denoted by V is the sum of the squares of the deviations of each measurement from the grand mean $\bar{x}$, i.e.

$$\text{Total variation} = V = \sum_{j,k}(x_{jk} - \bar{x})^2 \tag{3}$$

By writing the identity

$$x_{jk} - \bar{x} = (x_{jk} - \bar{x}_{j\cdot}) + (\bar{x}_{j\cdot} - \bar{x}) \tag{4}$$

then squaring and summing over j and k

We can have

$$\sum_{j,k}(x_{jk}-\bar{x})^2 = \sum_{j,k}(x_{jk} - \bar{x}_{j\cdot})^2 + \sum_{j,k}(\bar{x}_{j\cdot} - x)^2 \tag{5}$$

$$\bullet \quad \sum_{j,k}(x_{jk} - \bar{x})^2 = \sum_{j,k}(x_{jk} - \bar{x}_{j\cdot})^2 + b\sum_{j}(\bar{x}_{j\cdot} - x)^2 \tag{6}$$

We call the first summation on the right of (5) or (6) the variation within treatments since it involves the squares of the deviations of x_{jk} from the treatment means, $\bar{x}_{j.}$ and denote it by V_w Thus

$$V_w = \sum_{j,k} (x_{jk} - \bar{x}_{j.})^2 \tag{7}$$

The second summation on the right of (5) or (6) the variation within treatments since it involves the squares of the deviations of the various treatment means $\bar{x}_{j.}$ from the grand mean x and is denoted by V_b Thus

$$V_b = \sum_{j,k} (\bar{x}_{j.} - \bar{x})^2 = b \sum_{j} (\bar{x}_{j.} - \bar{x})^2 \tag{8}$$

Equations (5) and (6) can thus be written as

$$V = V_w + V_b \tag{9}$$

Expected values of the variations

The between treatments variation V_b, the within treatments variation V_w and the total variation V are random variables which assume the values V_b, V_w and V as defined in (8),(7) and (3).

$$E(V_b) = (a-1)\,\sigma^2 + b \sum_{j} \alpha^2_j \tag{10}$$

$$E(V_w) = a\,(b-1)\,\sigma^2 \qquad \text{where } \sum_{j} \alpha^2_j = 0 \tag{11}$$

$$E(V) = (ab-1)\,\sigma^2 + b \sum \alpha^2_j \tag{12}$$

From (11) if follows that

$$E\left(\frac{V_w}{a(b-1)}\right) = \sigma^2 \tag{13}$$

So that $\hat{S}^2_w = \dfrac{V_w}{a(b-1)}$ (14)

is always a best estimate (unbiased) of 2 regardless of whether H_o is true or not from

(10) and (11) only if Ho is true we have

$$E\left(\frac{V_b}{a-1}\right) = \sigma^2 \qquad E\left(\frac{V}{ab-1}\right) = \sigma^2 \tag{15}$$

In such case

$$\hat{S}^2_b = \frac{V_b}{a-1} \qquad \hat{S}^2 = \frac{V}{ab-1} \tag{16}$$

which provides unbiased estimates of σ^2.

If H_o is not true, then we have from (10)

$$E(\hat{S}^2_b) = \sigma^2 + \frac{b}{a-1} \sum_j \alpha^2_j \tag{17}$$

Analysis of variance table

We could compute V and V_b and then compute $V_w = V - V_b$. It should be noted that the degrees of freedom for the total

variation, i.e. ab-1, is equal to the sum of freedom for the between-treatments and within-treatments variations.

The hypothesis Test

Given the summary statistics, the calculations of the hypothesis test can be shown in tabular form.

Table 1.20 (The Hypothesis Test Table)

Variation	DF	Mean Square (MS)	F
Between treatments $V_b = b\Sigma(\bar{x}_{j.} - \bar{x})^2$	a-1	$\hat{S}_b^2 = \dfrac{V_b}{a-1}$	$\dfrac{\hat{S}_b^2}{\hat{S}_w^2}$ with
Within treatments $V_w = V - V_b$	a(b-1)	$\hat{S}_w^2 = \dfrac{V_w}{a(b-1)}$	a-1, a (b-1) df
Total, $V = v_b + v_w$ $= \Sigma_{j,k}(x_{jk} - \bar{x})^2$	ab-1		

It follows that a good statistic for testing the hypothesis H_o is provided by:

$\hat{S}_b^2 \big/ \hat{S}_w^2$ .If this statistic is significantly large we conclude that there is a significant difference between treatment means and thus reject H_o. *Otherwise we can either accept H_o or reverse judgment* bending further analysis. That is

$F_c < F_T$ (accept H_o)

$F_c > F_T$ (Reject H_o)

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Regression analysis models have been used in many areas like in science, business and engineering. Regression makes us understand the relationship between dependent variable N_p and independent variable x. The random quantity N_p is a function of one or more independent variables x_1, x_2, - - -x_4 [2]. The *background on models reveals that they vary in their level of formality*, explicitness, richness in detail and relevance. Models have several functions in explaining phenomena, making predictions, decisions and communicating knowledge [6]. Studies involving multivariate approaches to meta-analysis are more difficult to apply and justify [5]. This model will be concerned with analysis of entry behavior which will enable educators focus on better grades in their KCSE which will form a background *that influences the learner's academic performance in college* [18].

The self-variables are directly associated with both expectations and academic performance and students may use self-concepts to interpret behavior which may serve as motivational forces towards behaviours and plans consistent with their meaning. He found that self-concept and achievement are mutually reinforcing *to the extent that a positive change in one facilitates a positive*

change in the other. A teacher's label on a student's capability brings a self fulfilling prophesy in the learners which has a diverse effect [20]. Those labeled bright believe that they know everything while those labeled poor lose hope of passing their subjects. When students view themselves as being incapable in a subject, they develop a negative attitude towards the subject and will most likely not do well. Their previous performance can play a role in shaping their study habits even at entry level to the end of the final examination [10].

Learners performance basically depends on attitude which they develop as they begin in form one which can be passed onto them by teachers, parents and peers. They can also develop their own attitude in an effort to adjust to the environment [8]. Attitudes, beliefs, feelings, thoughts and emotions can be modified by new experiences. A teacher can trigger effort in learners which is facilitated by the strategies put in place nowadays like extra tuition and remedial teaching [13]. He observed that individuals will change their attitude in order to conform to those that the environment holds as the norm. Teachers do not use student centered approaches but lack of experiments and practical modeling activities and lack of professional exposure articulates issues relating to teaching in secondary schools [12]. Many teachers attributes this performance as negative attitudes this performance as negative attitudes by the students as well as

missing link between primary and secondary. Poor performance in Kenya is due to poor teaching methods and acute shortage of textbooks which are used as many as six students would share one textbook in some schools making it impossible for them to complete their homework [7]. Poor performance is due to the difficult language used in Mathematics classroom [19]. He said that words have a different meaning when used in common day English language compared with when used in Mathematics. To perform better, students need to understand the use of language in different subjects.

2.2 Statistical Model

Statistics is widely used in government business, natural science, social science and every area of serious scientific inquiry that must be subjected to statistical analysis for validation. Statistical modeling relies heavily on regression. Regression analysis is a strategic tool utilized by many of the world's top operations for marketing mix models, data mining and volume forecasting based on some set of parameters and initial conditions [3].

Sir Francis Galton (1892-1911) was the father of regression analysis after his experiments on sweet potatoes and hereditary patterns in heights of adult humans [1]. This project intends to involve a lot of graphs for easier interpretation by those in depth knowledge in statistics. Organizational activities play an

important role in the various conceptualizations of business models that have been proposed. Business models seek to explain both value creation and value capture [4].

Depending on what learners scored in KCPE the researcher was able to use regression analysis and one way Anova to determine the trend of performance in KCPE and KCSE to enable educators make viable decisions.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Research design, data and sampling

The study was carried out in Nyamira district, Nyamira County which comprised of 48 schools of which 15 schools were selected. These schools were classified as boys schools, girls schools and mixed schools. A sample of 15 schools was used to create a *sampling framework. This was part of the target population* which was used for the study. Simple random sampling was preferred for the study because it gave every school equal chances of being selected for the study [17]. Simple random sampling was used to select a sample of the accessible schools representing 31.25% of the population which is at least 30% of the population according to [14]. The first fifteen schools selected was used to represent the sample. The schools selected are National, *provincial and district schools.*

Official records of KCSE and KCPE results of these selected *schools were sought from the respective school administration by* document analysis. The researcher used document analysis to establish entry scores (x_1) and final grade (N_p) which was in the form of an equation as follows;

$$N_p = a + b_1 x_1 + b_2 x_2 + \text{ - - - } b_q x_q + \varepsilon$$

The dependant variable was N_p and the independent variables were $x_1, x_2 \text{ - - - } x_n$

3.2 Modeling the input and the output

Data was imported into stata version 12 (stata corp. USA) for analysis from excel file. Descriptive statistics including counts (with respective proportions) and means (with respective standard deviations) was used. For the Bivariate analysis the researcher used student t-test to assess the difference in KCSE scores among the male and female students. Univariate analysis was done using linear regression analysis where the significant factors were included into the multivariate linear regression model. Regression coefficients, respective 95% confidence intervals and p values were reported for each of the covariates fitted in the model. Adjusted R squared values were also used to assess the amount of variance accounted for by the covariates. Beta (standardized coefficients) was reported for histograms for the multivariate model to assess the importance of each of the covariates. Scatter plots, pie charts and tables were used to display the analysis results. We had $x_1 x_2 \text{ - - - } x_q$ as our q predictor variables related to the variable, final performance and a sample size of k=572. The linear relationship between the dependent (response variable) N_{pi} and one or more of the predictor variables then used a linear model to

relate N_p to the x's and were concerned with estimation and testing of the parameters in the model.

We can be able to expound two cases according to the number of variables.

- **Simple linear regression**

One N_p and one x for example if we wish to predict N_p based only on the number of variables required in the model.

- **Multiple linear regression (univariate multiple regression)**

One N_p and several predictor variables (x's).

We could attempt to improve our prediction of N_p by using more than one independent variable such as entry behaviour, end of form three exams, mock exams and gender

The classical linear regression model states that N_p is composed of a mean, M_o which depends in continuous manner on the independent variables x's and a random error ε

$$N_p = M_o + M_1 eb + M_2 ef3e + M_3 m_c + M_4 g + \varepsilon$$

With $K = 572$ independent observations on N_{pi} and associated values of the independent variables which reads to the complete model in a matrix form shown below.

$$
\begin{pmatrix}
N_p1 \\
N_p2 \\
\vdots \\
\\
\\
\\
N_Pk
\end{pmatrix}
=
\begin{pmatrix}
M_o + M_1\,eb11 + \text{- - -}\ M_4g14 + \varepsilon_1 \\
M_o + M_1\,eb21 + \text{- - -}\ M4g24 + \varepsilon_2 \\
\vdots \qquad \vdots \qquad\qquad \vdots \qquad \vdots \\
\\
M_o + M_1\,ebk1 + \text{- - -}\ M_4gk4 + \varepsilon_k
\end{pmatrix}
\qquad (3.1.1)
$$

That is

$$
\begin{pmatrix}
N_p1 \\
N_p2 \\
\vdots \\
\\
\\
\\
N_Pk
\end{pmatrix}
=
\begin{pmatrix}
M_o + M_1\,eb11 + \text{- - -}\ M_4g14 + \varepsilon_1 \\
M_o + M_1\,eb21 + \text{- - -}\ M_4g24 + \varepsilon_2 \\
\vdots \qquad \vdots \qquad\qquad \vdots \qquad \vdots \\
\\
M_o + M_1ebk1 + \text{- - -}\ M_4gk4 + \varepsilon_k
\end{pmatrix}
\qquad (3.2.1)
$$

Where the error terms ε_i's was assumed to have the following properties.

- The error term ε is a random variable with the mean or expected value of zero, that is
  $E\,(\varepsilon_j) = 0$

The model is linear and no additional terms are needed to predict N_p, all the remaining variation in N_p is purely random and unpredictable thus if $E\,(\varepsilon_i) = 0$, then

22

$E(N_{pi}) = M_0 + M_1 ebi1 + M_2 ef3ei2 + - - - + M_4 gi4$

and the mean of N_p is expressed in terms of these 4 predictor variables with no others needed.

- $Var(\varepsilon_i) = \sigma^2$ (constant)

The variance of each (ε_i) is the same for all values of predictors eb, ef3e, me and g which applies that $Var(N_{pi}) = \sigma^2$ and is also the same for all values of predictors eb, ef3e, me and g.

- $Cov(\varepsilon_i \varepsilon_i) = 0, i \# j$

The error terms are uncorrelated, from which it follows that N_p's are also uncorrelated that is $Cov(N_{pi}, N_{pi}) = 0$.

The values of ε for a particular set of values for the independent variables are not related to the value of ε for any other set of values.

- The error term is a normally distributed random variable reflecting the deviation between the N_p value and the expected value of N_p is given by

$M_o + M_1eb + M_2ef3e + M_3Me + M_4g$

M_o, M_1 - - - M_4 are constants for the given values of eb, *ef3e, - - - g, the dependant variable N_p is also normally* distributed random variable.

Assumptions of N_p:

- $E(N_{pi}) = M_o + M_1ebi1 + M_2ef3ei2 + - - - + M_4i4 = 572$
- $Var(N_{pi}) = \sigma^2 \ i = 1 - - - 572$
- $Cov(N_{pi} \ N_{pj}) = 0 \quad i \# j$

CHAPTER 4

## DATA ANALYSIS AND MODEL FITTING

4.1 Descriptive Statistics

The mean of female and male was found from the data of 572 candidates of which 348 were female and 224 were male. The *mean and standard deviation of KCPE, end of form three, mock and KCSE* was also calculated. They are tabulated as the table below (Table 4.10). The results are also displayed in histograms and pie charts which shows the behaviour of the curves.

Over half of the students, 348 (60.8%), was composed of females and 224(39.2%) were males. The mean score at KCPE was 41.26 (SD= 7.87), end of form 3 exam, 47.8 (SD=15.52), which had a higher mean score than the mock exams, 40.77 (SD=12.93). The *mean KCSE score, 51.25 (SD=16.32) was higher than* the end of form three and mock exams (**Table 4.10**).

Table 4.10: Summary of Covariates

Variable	N=572
	n(%) or Mean(SD)
Gender	
Female	348 (60.8)
Male	224 (39.2)
Scores	
KCPE	41.26 (7.87)
END OF FIII EXAM	47.8 (15.52)
MOCK	40.77 (12.93)
KCSE	51.25 (16.32)

Fig. 4.0: Gender Distribution

Gender distribution

39.16%

60.84%

Females   Males
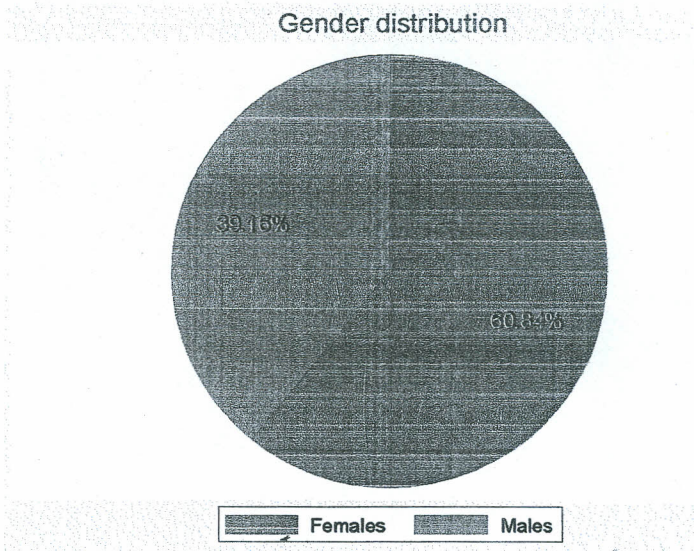
Histograms of scores distributions

Fig. 4.1: KCPE score distribution

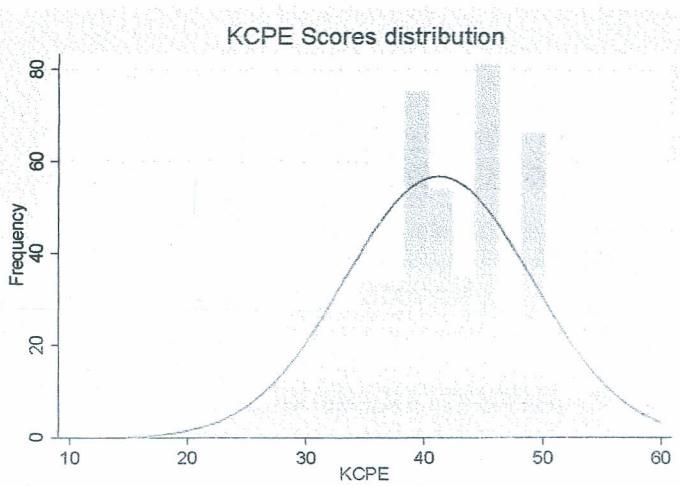KCPE Scores distribution

Fig. 4.2: End of Form III exam score distribution

ENDOFFIIIEXAM Scores distribution
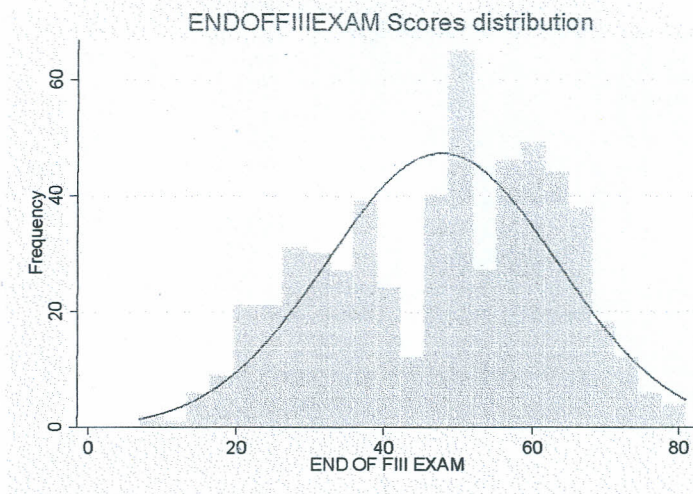
Fig. 4.3: Mock score distribution
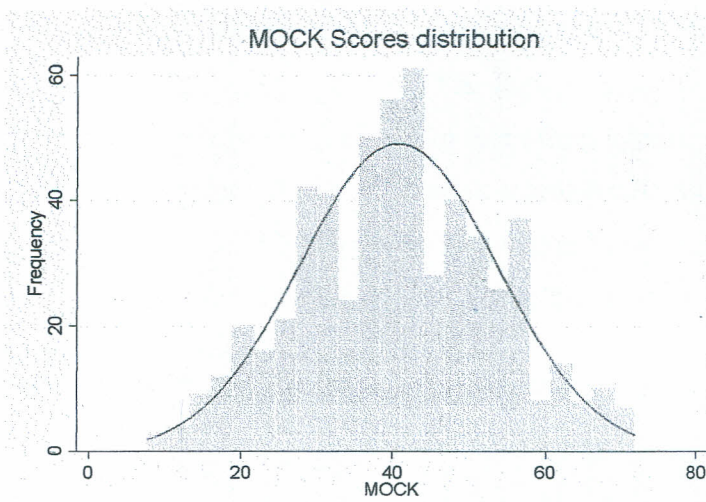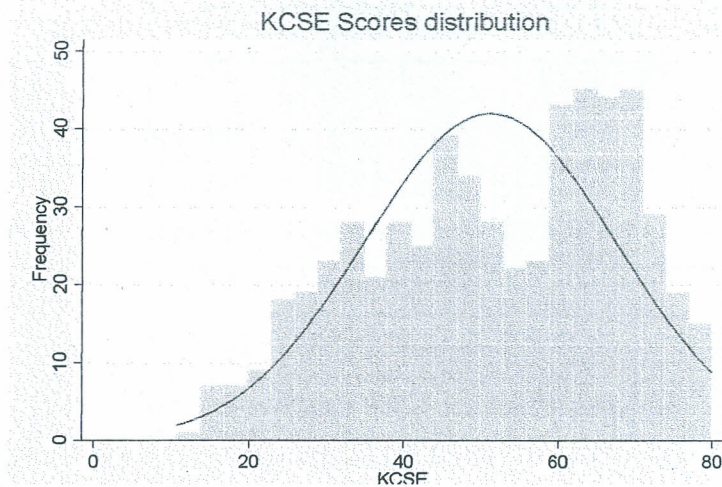

MOCK Scores distribution

Fig. 4.4: KCSE Scores distribution

4.2 Correlation Results

There was a significant positive linear relationship between KCPE performance and end form three exam (r=0.6512) Mock (r=0.5851) and KCSE (r=0.6614) scores. The highest positive correlation were between KCSE and end of form three exam (r=0.8960) and also between KCSE and mock scores (r=0.7397). The positive linear relationship between end of from three exam and mock was also high (r=0.7309), (Table 4.20). The scatter plots have been used to display this information.

Table 4.20 (Correlation table)

	KCPE	End of FIII Exam	MOCK	KCSE
KCPE	1			
End of FIII Exam	0.6512*	1		
MOCK	0.5851*	0.7309*	1	
KCSE	0.6614*	0.8960*	0.7397*	1

* p value <0.001

4.3 Scatter plots of scores Vs KCSE Scores

This information was well illustrated using scatter plots. The combinations are between KCPE and KCSE, end of form three and KCSE and mock. Since the variable KCSE was our main interest which was plotted against the rest.

Fig. 4.5: KCPE Vs KCSE Scores

KCPE vs KCSE scores

r = 0.6614   p <0.001
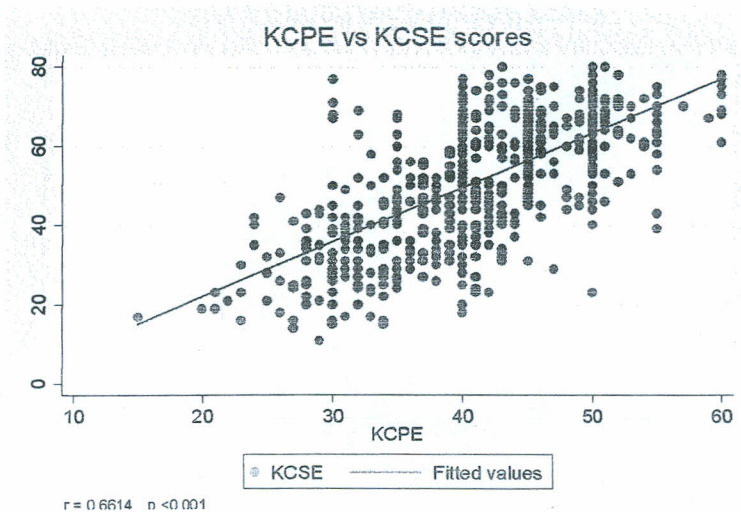
Fig. 4.6 End of Form III Vs KCSE Scores

ENDOF FIII EXAM vs KCSE scores

● KCSE ——— Fitted values

r = 0.8960 , p <0.001

Fig. 4.7 Mock Vs KCSE scores

ENDOF FIII EXAM vs KCSE scores

● KCSE ——— Fitted values

r = 0.8960 , p <0.001

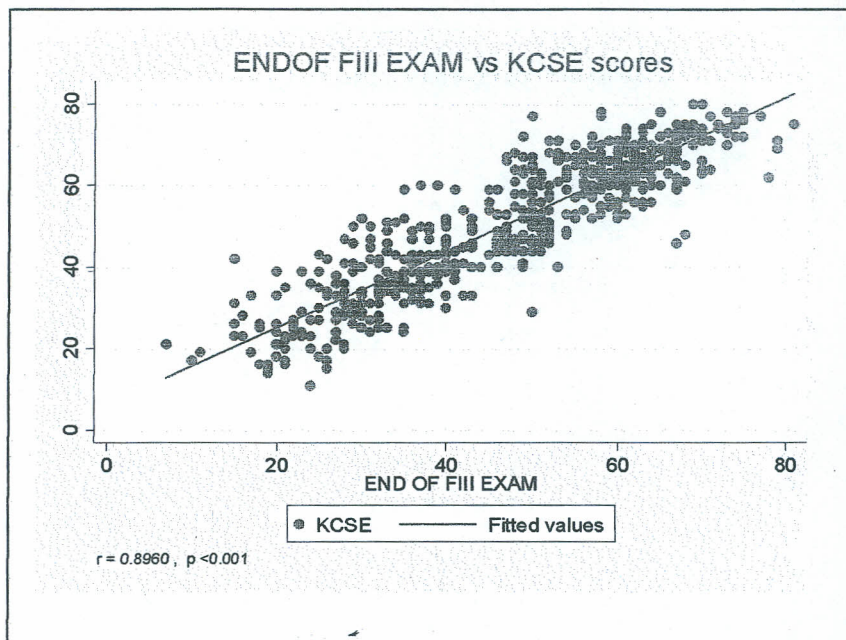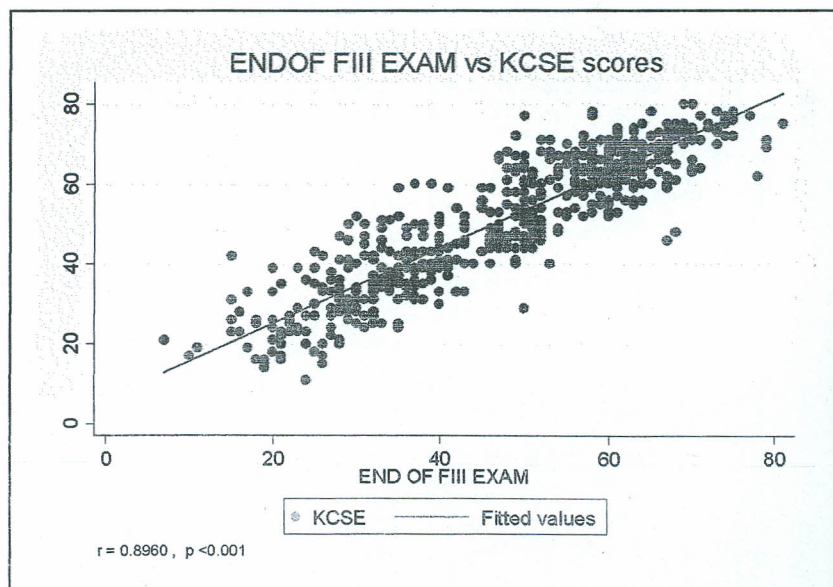## 4.4: Bivariate Analysis

There was significantly higher mean score at KCSE among females 55 (SD = 16.23) than males 45.43 (SD=14.71), p value <0.001.

Table 4.30 KCSE Score by Gender

Gender	n	KCSE Mean (SD)	Test Statistic	P value
Female	348	55.00 (16.23)	t-test	<0.001
Male	224	45.43 (14.71)		
Total	572	51.25 (16.32)		

* Statistically significant

4.5 Univariate Regression Analysis

The univariate regression analysis results indicated that for every score increase in KCPE there was a corresponding significant increase of 1.4 units (95% CI = 1.24, 1.50) at KCSE, P<0.001. The adjusted R^2 was fairly high; approximately 44% of the variance was accounted for by KCPE in the model.

End of form three exams was the strongest predictor accounting for about 80% of the variation in the model. The results showed that for every unit increase in end of form three score the students had a score increase of 0.94 units (95% CI = 0.90,0.98) score at KCSE, P<0.001. The mock score was an equally important predictor of the outcome, accounting for 55% of the variance. For

each unit increase in mock score there was a corresponding 0.93 unit (95%CI = 0.86, 1.00) increase in the KCSE score, P<0.001. In terms of gender, the male students compared to female had a significantly lower KCSE score of -9.57 units (95% CI:-12.21,-6.94), P<0.001.

Table 4.40 Univariate Linear Regression Model

KCSE Score	Adjusted R^2	Coefficient (95% CI)	Std. Error	t Value	P Value
KCPE Score	0.44	1.37 (1.24 1.50)	0.07	21.05	<0.001*
End of FIII Exam	0.80	0.94 (0.90 0.98)	0.02	48.17	<0.001*
MOCK Score	0.55	0.93 (0.86 1.00)	0.04	26.22	<0.001*
Gender [Ref: Female]	0.08				
Male		-9.57 (-12.21 -6.94)	1.34	-7.14	<0.001*

* Statistically significant

4.6 Multivariate Regression Analysis

All the significant covariates in the univariate model were included into the multivariate model. The model accounted for about 84% of the variance in KCSE scores. The standardized coefficient (Beta) indicates the most important predictors (have high absolute beta scores) of KCSE score as end of form three exam, mock, gender then finally KCPE score. The results show that for each unit increase in end of form three exam score

holding other factors constant the students fared better at KCSE by 0.69 units (95% CI = 0.63, 0.75), P<0.001.

Similarly, for every unit increase in KCPE exam score holding the other factors constant there was a corresponding increase of 0.20 units (95% CI= 0.11, 0.30), P<0.001. Taking mock exam score into consideration, there was a corresponding increase in KCSE exam score of 0.26 units (95% CI=0.19, 0.32) for each unit increase in mock score, P<0.001. The male students fared worse than the female students holding other factors constant by about four unit scores, -3.99 units (95% CI = -5.17, -2.80), P<0.001.

Table 4.50 Multivariate Linear Regression Model

KCSE	Adjusted R²	Coefficient (95%CI)	Std Error	t value	P value	Beta
	0.84					
KCPE		0.20(0.11,0.30)	0.05	4.31	<0.001*	0.0987
End of FIII		0.69(0.63,0.75)	0.03	23.05	<0.001*	0.6555
Mock		0.26(0.19,0.32)	0.03	7.72	<0.001*	0.2934
Gender Ref: Female						
Male		-3.99(-5.17,2.80)	0.60	-6.61	<0.001*	-0.1195
Constant		-1.06(-1.96,4.08)	1.54	0.69	0.493	
*statistically significant						

Table 4.50: Multivariate linear Regression Model

4.0.7 The Fitted Full Model

From table 4.50 above our model is formulated as:

$$N_p = M_0 + M_1 (eb) + M_2 (ef3e) + M_3 (me) + M_4 (g) + \varepsilon$$

$$= 1.06 + 0.20 (eb) + 0.69 (ef3e) + 0.26 (me) + (-3.99) (g) + \varepsilon$$

Table 4.60: Table of Means

SCHOOL	KCPE	KCSE	TOTAL (T)	T2
A	45.6	64.3	109.9	12,078.01
B	41.3	50.4	91.4	8,353.96
C	40.3	44.2	84.5	7,140.25
D	37.4	35.0	72.4	5,745.64
E	37.3	29.2	66.5	8,649
F	32.3	43.5	75.8	7,569
G	43.1	48.9	93	8,649
H	38.8	48.2	87.0	7,569
I	38.6	40.4	79.3	6,288.49
J	40.2	47.2	87.4	7638.76
K	47.7	71.9	119.6	14,304.16

L   46.6   65.7   112.3   12611.29

M   42.6   64.0   106.6   11363.56

N   41.7   58.8   100.5   10,100.25

O   44.8   66.7   115.5   12,432.25

Total 618.3 778.4 1231.4 66969.315

$K = 2, n = 15$

Correlation term, $C = \dfrac{G^2}{kn} = \dfrac{1396.7^2}{30} = 65025.696$

Sum squares $= \sum\sum y.^2{}_j = 68378.69$

Total sum squares $= \sum\sum y.^2{}_j - \dfrac{G^2}{Kn} = \dfrac{68378.69}{30} - 65025.696$

$$= 3352.99$$

S.S Between groups $= 618.3^2 + 778.4^2 - \dfrac{2737.2^2}{30}$

$$= 854.4$$

S.S within groups (Error) $= 3352.99.69 - 854.4 = 2498.59$

Table 4.70 Anova Table

Source of variation	S.S	D.F	M.S.S	F
Between groups	854.4	1	854.4	$\dfrac{854.4}{89.235} = 9.57$
Within groups	2498.59	28	89.235	
Total	3352.99	29		

$F_C\ 0.95 = 9.574 > F_T = 4.\ 2.$ at $(1, 28)$

We can reject the hypothesis and conclude that there was no difference between the means of KCPE and KCSE at 5% confidence level.

$F_{C\ 0.99} = 9.574\ F_{T\ 0.99} = 7.63$ at $(1, 28)$ df

There was no difference between the means of KCPE and KCSE at 1% confidence level.

CHAPTER 5

Summary, Conclusion AND RECOMMENDATIONS

The mean for KCSE was, 51.25 which was higher than end of form three mean, 47.8 which was higher than KCPE exam mean, 41.26. Mock results recorded the lowest mean, 40.77. Girls were more (60.8%) compared to boys (39.2%) (**Table 4.10**).

The standard deviation for KCSE was highest, 16.32 followed by end of form III 15.52 then mock 12.93 and lastly KCPE, 7.87. The histogram for KCPE scores displayed a distribution which his concentrated around the centre while KCSE and end of form III displayed a normal distribution which is slightly skewed to the right. This implies that KCSE and end of form III were better performed than KCPE and mock.

Correlation between end form III and KCSE was highest (r=0.7397) then the correlation of KCSE and mock was the least (r=0.6614) (**Table 4.20**). This implies that what the learners scored at end of form III was a clear reflection of what they scored at KCSE. It showed that what learners scored at KCPE can improve if a lot of effort is put by the learners, parents and teachers in guiding them. Mocks showed clearly what the learners were likely to score as their minimum since it was closer to KCSE.

There was a strong relationship between KCSE and end of form III as it was seen by the scatter diagram. The results showed that for every score increase in KCPE there was a corresponding significant increase of 1.4 unit (95% CI=1.24, 1.50) at KCSE. About 44% of the variance was accounted for by KCPE in the model. *End of form III exams was the strongest predictor* accounting for about 80% of the variation in the model. Also mock exams accounted for about 55% of the variation in the model. Mock exams were equally important predictor of the outcome since for each unit score, there was a corresponding 0.93 unit increase in KCSE and for every unit increase in end of form III exams the students had a score increase of 0.94 units. Male students compared to female had a lower KCSE score of -9.57 units (**Table 4.40**).

The univariate model was included in the multivariate model *which accounted for about 84% of the variance in KCSE scores.* The Beta (the standardized coefficient) indicated that end of form III has a high beta score which showed that it was the highest predictor of KCSE followed by mock, gender then KCPE score (**Table 4.50**). The results showed that every increase in end of form III scores holding other factors constant there was a corresponding increase of 0.20 units and mock exam scores holding other factors constant, there was an increase of 0.26 units. Male students fared worst than female holding other factors

constant by -3.99units. Male students fared worst than female holding other factors constant by -3.99 units.

Using a one way Anova at 5% confidence level, we can reject the hypothesis and conclude that there was no difference between the means i.e.

$F_{C0.95} = 9.918 > F_t = 2.77$ (**Table 4.70**) Also at 1% confidence level there was no difference between the means i.e.

$F_{C0.99} = 9.918 > F_t = 4.16$

Anova for means of schools showed that we can reject the null hypothesis of equal means at 0.05 and 0.01 confidence levels

i.e. $F_{C0.95} = 5.914 > F_{t0.95} = 1.92$ $F_{C0.99} = 5.914 > F_{t0.99} = 2.5$

(**Table 4.90**)

Looking at the analysis, we conclude that end of form III exams were a *clear reflection of KCSE. This implied that they had* thoroughly been prepared for three years unlike when they did mock and the duration was shorter that is three months to KCSE. Mock exams were challenging and were not a good measure of KCSE that was why learners were young and not yet realized their dreams. Using the Anova tables, schools selected had no differences in their means at 0.01 and 0.05 confidence levels. Also we could reject the hypothesis of equal means at 0.01 and 0.05 *confidence levels. This implies that entry behavior was not a good*

predictor of the final performance and therefore there are better predictors of KCSE like end of form III and mock exams etc.

We recommend that there are other better measures of KCSE other than KCPE. I also recommend that any student who has done KCPE can pass KCSE if he/she works hard. Finally the Ministry of Education should encourage those who had low marks at KCPE that they can proceed to Form One and work hard and pass KCSE.

References

[1] Allen Webster: Applied statistics for Business & Economics an *Essential version* (Third Edition) McGraw Hill, Singapore (2005)

[2] Alvin C. Methods of multivariate Analysis. (Second edition): Wisley, Canada (2002).

[3] Amy Dahan Dalmedico: History and Epistemology of models on meterology, (1963).

[4] Anderson T.W.: An Introduction to multivariate statistical Analysis (Third Edition) Willy, New York, (2004).

[5] Atkinson T.W. An Introduction to Multivariate Statistical Analysis (Third Edition) Willy, New York, (2004).

[6] David J.H.K. Anatomy of Business Models. A syntactical Review and Research Agenda London Business School, (2010).

[7] Eshiwani G: Mathematics Education around the world Bridging policy and Practice: The role of Algebra in the middle and secondary Mathematics Curriculum. A paper presented at an

international seminar in Park City. Mathematics Institute Utah, USA, (2001).

[8] Evans, K.M: Attitudes and interest in Education. Routledge and Kegan Paul, London, (1972).

[9] Hamachek D.E: Encounters with the self Holt, Rinehart and Winston, New York, (1971).

[10] Maritin E.K: The academic self-concept and the teachers' perception. Their relationship to pupils' grade attainment in rural Kenya. Harvard University, London, (1997).

[11]   McCcllock : Testing for competence rather than intelligence. American psychologist, USA, (2003).

[12] Miheso K.M: Factors affecting Mathematics among secondary school students in Nairobi province; Kenya, unpublished PhD Thesis, KU, (2012).

[13] Muasya, I.W: Students' self concept of academic ability and students' academic Performance in secondary schools. Kenyatta University, Nairobi, (1992).

[14] Mugenda and Mugenda :Qualitative and quantitative approaches in Research United Nations Avenue, Gigiri, (2003).

[15] Murray R.S: Probability & Statistics. MC Graw Hill, Inc, (1975)

[16] Nash, R: Teacher's Expectations and pupil learning. Routledge and Kegan Paul, London, (1976)

[17] Oso, W.Y. and Onen, D: A general guide to writing research proposal and report. A handbook for beginning research. Nairobi options press and publishers, (2005)

[18] Reitzes, D.C. and E. Mutran: Significant others, and self-conceptions. Social Edu, (1990)

[19] Wasike D.W: Effects of socialized Mathematical language module on students' Understanding of Mathematics, (2003)

[20] Wawire, V: Perceptions of students and teachers on the impact of streaming on student's self-image on academic performance, Kenyatta University: Nairobi,(1996)

[21] Weber, Eric: The indispensable Employee: A survival manual, Berkel New York, (194