University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Mathematics

Application of Ordinal Logistic Regression in Analyzing Students' Performance at Kenya Certificate of Secondary Education Level in Kiambu County.

Research Report in Mathematics, Number 10, 2018

Kenneth Benson Muya

August 2018

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Application of Ordinal Logistic Regression in Analyzing Students' Performance at Kenya Certificate of Secondary Education Level in Kiambu County.

Research Report in Mathematics, Number 10, 2018

Kenneth Benson Muya

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master of Science Project
Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Prepared for The Director
Graduate School
University of Nairobi

Monitored by School of Mathematics

Abstract

Background: In the Kenyan education system (8-4-4 system), education progresses from pre-school to primary school to secondary school to tertiary education. Transition from one stage/ level to another is normally done after evaluation. Examinations are done at the end of each stage, National examinations are used to evaluate students for transition from one level to another

Movement from primary level to secondary level is determined by good performance in the KCPE examination. The transition from secondary level to tertiary level is determined by good performance in the KCSE examination. These examinations are normally set, administered and evaluated by the Kenya National Examinations council (KNEC), a body set aside by the kenyan government for this purpose.

Method: This study aimed at analyzing the overall performance of students at KCSE level in Kiambu county. The subjects used in the study included Mathematics, English, Swahili, Biology & Chemistry. The students gender and also the type/category of secondary school attended by the students were also considered.

Ordinal logistic regression was the method used in analyzing the students performance in the year 2014 & 2015. Stratified random sampling technique was used, Kiambu county was stratified into the 12 different sub-counties then schools were selected randomly from them, Selection of schools was on the basis of school category.The samples for both the years comprised of approximately 24% of the entire population of candidates.

Results & Conclusion: After analysis , the findings showed that the subjects that contributed the most to the students overall performance in both the years were Swahili & Biology. Mathematics did not contribute much. Students gender did significantly have an effect on the students overall grade.

Keywords: Ordinal logistic regression, KCSE, National Schools, Extra county schools, County Schools, Sub County schools.

Dissertations in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this project report is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

KENNETH BENSON MUYA
Reg No. I56/87473/2016

In my capacity as a supervisor of the candidate, I certify that this report has my approval for submission.

Signature Date

DR. IVIVI JOSEPH MWANIKI
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: jimwaniki@uonbi.ac.ke

Dedication.

I dedicate this project to my wife Maureen, Son Ian, Daughter Ivy , my Mother Rebecca, My sisters Edna, Carol and Tracy. I thank them all for their endless support and patience especially when i had to attend classes late in the evening.

May they all live to love, cherish and appreciate the power of education.

List of Abbreviations/Acronyms.

1. ANOVA – Analysis of Variance.
2. Bio – Biology Subject.
3. Chem – Chemistry Subject.
4. CDF – Constituency Development Fund.
5. Eng – English Subject.
6. Et al – et alia/ et alii – and others
7. E County – Extra County Schools.
8. H_0 – The Null Hypothesis.
9. H_1 – The Alternative Hypothesis.
10. KCPE – The Kenya Certificate of Primary Education.
11. KCSE – The Kenya Certificate of Secondary Education.
12. KLB – Kenya Literature Bureau.
13. KICD – Kenya Institute of Curriculum Development.
14. Kis – Swahili Subject.
15. KNEC – Kenya National Examination Council.
16. ln – Natural Logarithm.
17. Maths – Mathematics Subject.
18. MLE – Maximum Likelihood Estimation / Estimator.
19. MOEST – Ministry of Education Science and Technology.
20. NGO – Non - Governmental Organization.
21. OLR – Ordinary Logistic Regression.
22. OLS – Ordinary Least Squares.
23. OD – Odds Ratio.
24. Phy – Physics Subject.
25. PLS – Partial Least Squares.
26. POM – Proportional Odds Model.
27. SCHL CATE – School Category.
28. SMASSE – Strengthening of Mathematics and Sciences in Secondary School Education.
29. S County – Sub County Schools.
30. TSC – Teachers' Service Commission.
31. UNESCO – United Nations Educational, Scientific and Cultural Organization.
32. χ^2 – Chi– Squared.

Contents

Bibliography 64

List of Tables

Acknowledgments

I am extremely grateful to the almighty God for the opportunity, energy and wisdom he gave me to pursue the Masters Degree.

I am grateful to the University of Nairobi, School of Mathematics (som). Mainly my supervisor Dr. Ivivi Joseph Mwaniki for the assistance, continuous advice & guidance, invaluable support and encouragement throughout the entire project. He guided me, corrected me and gave me the freedom to effect corrections my own way.

I pass my appreciation to all the University of Nairobi, School of Mathematics (som) lecturers who took me through the whole course work. Prof. Patrick Weke (Director School of Mathematics), Dr. Nelson Owuor, Dr. George Muhua, Dr. J. Ndiritu, Dr. Vincent Oeba, Dr. Japheth Awiti, Mrs. Ann Wang'ombe, Dr. Jacob Ong'alo & Dr. Jared Ongaro.

I acknowledge KNEC through the CEO Ms. Mercy Gathigia Karogo, TSC kiambu county and sub county offices for the provision of necessary data for analysis.

I pass my regards to all my classmates (class of 2016-2018) for their tireless support, teamwork and encouragement to carry on to the end of the Masters program.

May God bless you all and your families. ....

Kenneth Benson Muya

Nairobi, 2018.

1 INTRODUCTION

1.1 BACKGROUND OF THE PROBLEM.

Education is one of the most priced and treasured investment in the whole world. Most developed and developing nations normally invest a lot in their education systems.Education is primarily perceived as a means of solving social problems.

As a means of ensuring that education, moreso basic education is provided, many developing and developed nations for instance as United States of America, Australia and the United Kingdom have put in place mechanisms to provide free elementary and secondary school education. Developing countries such as Kenya have also come in, through the MOEST to ensure the is provision of subsidized primary and secondary education in order to achieve the 2030 millennium goals.

Apart from just the provision of free primary and subsidized secondary education;the government has started up initiatives to expand the existing primary and secondary schools and also building new schools through the various 'Harambee' projects and funds from the constituency development funds so as to accommodate the increasing demand for education.

Education plays a vital role in career development and career shaping.

Students start shaping their careers as early as during elementary studies, that is during their primary education. Thus the provision of education at this early level or stage is very vital.Students in secondary school are in their exploration phase in life. This is a stage where they are likely to be developing their careers (Patton and McMahon, 2014).

According to researches done by UNESCO, Many developing and developed Countries allot a huge percent of their wealth to Education sectors,(UNESCO, 2005) leading to a considerable expansion in Educational undertakings world wide.
Kenya being a developing country has not been left behind, Statistics show that Kenya allocates a huge share of its resources to the sector UNESCO 2005.

Apart from resource allocation, the government has various initiatives that try to make schooling more effective and efficient to the learners, Such initiatives include the protection of the rights of children for example through the campaigns against female genital

mutilation (FGM) and protection against early marriages. The governments working with various Non governmental Organizations (NGO,s) Provide sanitary towels to students in need.Examples of such organizations are Crossroads Global Hand and Mfariji Africa.The government is also working on the school feeding programs to the needy students.

Since independence in 1963, Education was seen as a means of eradicating poverty. Therefore, the government introduced the 8-4-4 education system, the system has been providing a practical oriented curriculum that aims at providing employment opportunities for those who go through it (Eshiwani,1993).

Despite all these provision and support that the education sector is getting. It faces various challenges, that is poor overall performance of students at KCSE level.

The body that is entrusted with the duty of assessing students at the end of the form 4 level that is the Kenya National Examination Council (KNEC), has registered poor overall performance over the years and more specifically in the mathematics and science examinations. This is shown in the table below (Table 1.2).

Table 1.1. Performance of Mathematics and Science Subjects in KCSE Compared by Gender Between the year 2002 and 2010

YEAR	GENDER	MATHEMATICS	PHYSICS	CHEMISTRY	BIOLOGY
2002	FEMALE	16.44	26.61	22.05	24.58
	MALE	22.53	30.89	26.62	28.34
2003	FEMALE	16.05	29.07	24.04	27.23
	MALE	22.10	32.28	29.30	31.35
2004	FEMALE	15.39	31.41	25.79	32.91
	MALE	21.34	35.25	30.43	37.64
2005	FEMALE	12.97	32.85	24.54	27.24
	MALE	18.49	35.99	29.44	30.01
2006	FEMALE	15.75	39.07	22.56	25.00
	MALE	21.87	40.42	27.01	29.84
2007	FEMALE	15.74	39.04	22.65	38.99
	MALE	25.10	42.23	27.69	44.70
2008	FEMALE	17.71	26.32	20.93	28.49
	MALE	24.31	26.95	24.27	32.01
2009	FEMALE	18.11	29.93	17.56	25.15
	MALE	23.63	31.88	20.43	29.08
2010	FEMALE	19.71	33.46	22.80	26.99
	MALE	25.75	35.76	26.62	31.24

Table 1.2. Table on the national mean performance of different subjects at KCSE between the years 2010 and 2015.

SUBJECT	OVERALL MEAN AS A PERCENTAGE					
	2010	2011	2012	2013	2014	2015
ENGLISH	39.26	36.74	38.13	35.23	47.68	40.28
KISWAHILI	44.34	49.01	36.32	39.91	47.08	47.88
MATHEMATICS	19.17	21.00	25.30	25.10	24.02	26.88
BIOLOGY	26.71	31.72	25.38	28.70	31.83	34.80
PHYSICS	31.50	32.94	32.53	36.87	38.84	43.68
CHEMISTRY	22.89	23.40	27.72	25.45	32.16	34.36
HISTORY	41.73	38.45	37.14	41.78	53.83	51.71
GEOGRAPHY	33.86	38.15	43.09	41.02	44.04	43.92
C.R.E	46.05	49.38	44.34	51.93	53.15	52.48
AGRICULTURE	31.25	34.26	32.03	31.94	40.82	44.81
B/STUDIES	37.28	42.61	51.00	53.64	46.82	43.76

source 2010-2016 KCSE examination report

1.2 STATEMENT OF THE PROBLEM.

Even though sciences (Biology,Chemistry and Physics) & Mathematics subjects play a great role in preparing students for careers that are science based. Careers that will help the country meet its 2030 development goal. The performance on these subjects tends to be very poor. The failure rate in Mathematics and Science (Biology,Chemistry and Physics) subjects is well above 50%.

It has been noted that learners perform poorly in Mathematics and science(Biology,Chemistry and Physics) subjects due to poor understanding,comprehension and application skills and this is attributed to poor performance in Language subjects that is English and Swahili.

There are also aspects of gender disparity when it comes to the performance of different subjects. Different school categories also have different level of performance.

Though it is a national issue, the study focuses on Kiambu County.

1.3 OBJECTIVES OF THE STUDY.

1.3.1 MAIN OBJECTIVE.

Analyzing the performance of students at secondary school level using past KCSE results.

1.3.2 SPECIFIC OBJECTIVES.

(a.) To determine the subject that contributes the most to students overall performance at KCSE level.

(b.) To establish if there exists a relationship between students gender and overall KCSE performance.

(c.) To establish if there exists a relationship between the category of school attended by a student and the overall KCSE performance.

1.4 RESEARCH QUESTIONS.

(a.) Which subjects contribute the most to the overall performance of students in secondary school at KCSE level?

(b.) Does the gender of students influence their overall performance at KCSE level?

(c.) Is there a relationship between the type/category of school attended by a student and the overall KCSE performance?

1.5 RESEARCH HYPOTHESES.

1. H_{01} –Sciences and Mathematics do not contribute the most to the students mean grades.
   H_{11} –Sciences and Mathematics contribute the most to the students mean grades.

2. H_{02} –There is no association between students gender and their mean performance at KCSE level.
   H_{12} –There is an association between students gender and their mean performance at KCSE level.

3. H_{03} –There is no difference between the performance of students in the different categories of schools.
   H_{13} –There is a difference between the performance of students in the different categories of schools

1.6 JUSTIFICATION/SIGNIFICANCE OF THE STUDY.

Despite the fact that many studies have been done to analyze the performance of secondary school students at KCSE level,very few studies have been done to analyze the effects of the performance on individual subjects i.e Mathematics, English, Swahili, Biology, and Chemistry on the overall performance of students at KCSE level.
The reason for the choice of just 5 subjects to analyze the effect of individual subjects on the overall performance of students at secondary school level is due to the fact that:

(i). English, Kiswahili and Mathematics are compulsory subjects categorized as group 1 subjects.

(ii). Chemistry and Biology are compulsory science subjects in most secondary schools and they are done by most students.

1.7 SCOPE OF THE STUDY.

This study was carried out using data from past KCSE examinations for selected schools within Kiambu county, the schools are to be categorized as National schools,County schools, sub-county schools and district schools.Students are to be further classified in terms of gender as either female or male. The selection will also be based on the schools population in terms of candidates who sat for the national examination (KCSE) in the selected year.

1.8 RESEARCH ASSUMPTIONS.

The following assumptions were considered:

(a). Data analysis was done based on KCSE 2015 results which is a standardized examination which is used to assess learners at the end of there four year secondary education period.

(b). All selected schools have teachers appointed by the Kenyas' teacher employing body that is the Teacher's service commission (TSC).

(c). All students in the schools under the study have equal study time.

(d). All students in the schools under the study subjected to the same academic syllabus/curriculum.

(e). Students had similar learning backgrounds in primary schools and any differences in learning, was as a result of classroom experiences in secondary schools.

(f). All the schools in the study use the same syllabus books.

(g). Teaching and learning is currently going on in the schools being used in the study.

1.9 STUDY LIMITATIONS.

1. The researcher would have wished to incorporate several counties in the study but due to logistics, distance, time factor, financial constraints and availability of data. It was a challenge.
2. The sample for the research was be picked from Kiambu county only and not from the entire country. This implies that the sample may not represent all high schools in Kenya.

2 LITERATURE REVIEW

2.1 INTRODUCTION.

The chapter contains related reviewed published writings . It focuses on studies done by other researchers, the methods they used and their findings.

2.2 SECONDARY SCHOOL EDUCATION IN KENYA.

Secondary education in Kenya begins immediately after primary school. It mainly caters for learners between the ages of 13 to 19 years. Secondary schools in Kenya are categorized as National, Extra county, county or sub-county schools for the public schools, we also have private schools. These categories are further sub divided according to gender; that is male only, female only and mixed schools. We also have boarding and day schools Kremer (2009).
The public schools are partially sponsored by the government and hence their school fees is subsidized.

National schools receive students from the whole country. County schools admit most of their students from within their respective counties. Extra county schools mainly admit students within the county of location and neighboring counties. Sub county schools admit most of its students from within their sub counties while day schools admit students from the immediate neighborhood.

Private schools which are owned by individuals and organizations. They have their management decide on the mode of admission.

In the recent past, due to the increase in the number of students clearing standard eight (8), that is the highest level of primary education. The secondary schools have been forced to expand in order to accommodate the increasing number of students joining secondary school.

A report by the world bank (2008) observes that out of 7.6 million Kenyans who attended primary school it is only 810,000 or 0.81 million who get admitted to secondary schools. The number even reducing further for those who manage to join universities to around 10,200.Okenyo,(2010).

A study conducted by Asena (2016) on the effect of subsidized free day secondary education in ensuring students are retained in schools. The aim of the study was to determine how free day secondary education contributes to students being retained in secondary schools in Kenya.

Asenas' objective was to study how availability of finances and teaching resources contributes positively to students understanding in school.

Population targeted was made up of 3993 stakeholders in the department of education in Bungoma county. Asena sampled 340 respondents.

Data was analyzed by use of descriptive statistics and content analysis.

Asena's study showed that the number of students in high schools was increasing drastically, this is attributed to the introduction of subsidized free day secondary education in the year 2008 by the government. It was aimed at providing basic education to all its citizens. The result of the subsidized free day secondary education is huge enrollment and transition rates that have contributed to shortage of teachers & high teacher student ratio. Asena et.al (2016).

2.3 ACADEMIC PERFORMANCE OF STUDENTS IN SECONDARY SCHOOL.

Many studies and researches have been done on the main causes of poor performance (Maiyo, 2009 and Orodho, 2004).

A study done by Cynthia (2014) on factors influencing the academic performance of students in KCSE Examination in Roysambu constituency in Nairobi.
The target population was 750 form 4 students and 145 teachers. Stratified sampling technique was made use of in the study. study objectives included:

1. Analyzing the influence of availability of learning resources on KCSE performance in Roysambu constituency.
2. To examine the influence of discipline on students performance in KCSE in Roysambu constituency.
3. To study the influence of home environment on students' performance.

The research used a descriptive cross-sectional survey design to explore the factors influencing students' performance in KCSE.

The research findings indicated that discipline is paramount in the performance of students Teaching methods also had a positive influence on the students performance. Schools with adequate teaching and learning resources posted better KCSE results than schools without adequate teaching and learning resources. Another finding is that home based factors also influenced students KCSE performance. These factors included: Parents involvement in learners school activities and the parents level of education (Cynthia 2014).

A research conducted by Samuel (2014) on the determinants of students performance in KCSE using ordinal logistic regression. Samuel's objectives included the following:

1. To find out the effects of private and public primary schooling on a students secondary school academic performance.
2. Identify the determinants of students' performance and achievements based on transition from different categories of schools, that is from private primary school to public secondary schools.

The target population was secondary schools in Kiambu county, Samuels' sample included 6 secondary schools in Kiambu county, data used was the KCSE results for the year 2013.

Samuels' methodology entailed exploratory and confirmatory analysis, he used ordinal logistic regression. The determinants of students performance at KCSE included the age of the learners, type of primary school attended (public or private), students gender and the KCPE marks attained.

Samuels' findings indicated that at primary school level, more students attended private schools than public schools while at secondary school level more students attended public secondary schools than private secondary schools.The table (Table 2.1) below shows the distribution of students in different categories of primary and secondary school, the categories are private and public schools.

Table 2.1. Distribution of students in different types of primary and secondary schools.

LEVEL / TYPE	*PRIMARY*	*SECONDARY*
PUBLIC SCHOOL	324	583
PRIVATE SCHOOL	351	92

Samuel's findings indicated that, more male students from the sample took the KCSE examination at secondary school level compared to female students. The percentage of male students was 59.7% while that of female students was 40.3%.

Samuel's findings further showed that the overall best performance was from students in National schools. There was also less movement from public primary schools to private secondary schools as compared to the movement from private primary schools to public secondary schools as indicated in table 2.1 above.

The study indicated that there were more female than male enrollment to private schools. The enrollment was 58.7% and 41.3% respectively. This being an indication that gender parity has not yet been achieved in school enrollment.

Students who attended public secondary schools achieved better results compared to those who attended private secondary schools. The study indicated that the significant factors in determining a students performance at KCSE were the KCPE marks obtained by the same students, the gender of the students and the category of secondary school attended by the students.

The primary school type & students age were not significant to the performance at KCSE level. Despite this fact an increase in age beyond 20 years had a negative effect on KCSE performance. A transition from private primary school to public secondary school significantly influenced students performance.

2.4 PERFORMANCE OF DIFFERENT STUDENTS GENDERS IN DIFFERENT GENDER CATEGORIZED SCHOOLS.

Several studies have been done to analyze the performance of students at secondary school level. among the studies done is one by Elvis K M (2015) who analysed the KCSE performance in Nakuru using Generalized estimating equations. His study showed that boys in boys schools scored better than boys in mixed gender schools. Girls in girls schools scored better than girls in mixed gender schools. In the mixed schools boys did better than girls.

The study showed that the grades of students in single gender schools that is Boys only and girls only schools were better than the grades of students in mixed schools.

A study conducted by Lydia (2013) on the factors that determine girls' scores in science, mathematics & technology subjects in public secondary schools. The research objective was to determine female students grades in science, mathematics & technology subjects in public secondary schools.

Lydia, used an ex-facto survey research design, Descriptive and Inferential statistics were derived. The research findings indicated that the qualification of teachers is a significant determinant in the learners performance. The study showed that girls performed poorly in science subjects compared to the other subjects.

2.5 ACADEMIC PERFORMANCE AT DIFFERENT SUBJECT LEVELS.

UNESCO, (2005) and Alidou, (2009), point out clearly that poor achievement from learners is not due to the students having inherent cognitive problems but rather, inadequate mastery of the language of instruction.

An analysis of past performances by KNEC and the Ministry of Education Science and Technology (MOEST), show key areas that candidates must prepare in and common mistakes that they should avoid as they sit for tests and examinations.

A research done by Mwiti (2016) on the grade scores in science subjects. The research objectives were:

1. To establish the relationship between achievement in languages, mathematics & sciences.
2. To model the performance in sciences given the grades in english, swahili & mathematics.

The data used for the study was KCSE results for the year 2014. The study population was 438660 candidates, while the target population was 65,535 candidates who took the three sciences (biology, chemistry & physics) at KCSE level.

Data analysis was conducted using partial least squares (PLS) regression to establish the relationship between mathematics, Languages & science subjects and predict the achievement in science subjects given the grades scored in mathematics and languages.

Analysis indicated that there existed a correlation between english, swahili and mathematics and that the performance in physics and chemistry is mainly influenced by that of mathematics. More findings were that gender negatively correlated to the achievement in sciences. The type of school attended being positively correlated to the science performance, where by national schools posted the best results followed by extra county schools then the county schools.

2.5.1 PERFORMANCE IN BIOLOGY.

Biology has not been performed very well nationally. The tables bellow (table 2.2 & 2.3) obtained from the Kenya National Examination Centre (KNEC 2010, report) show that there is a constant poor performance in the science (Biology) subject.

Table 2.2. National Percentage passes in Biology Between the year 2004 and 2010 according to KNEC 2010 Report

SN	YEAR	% of students with B+ and above	% of students with D and below
1	2004	12.03	36.67
2	2005	7.70	43.61
3	2006	6.13	49.64
4	2007	8.79	40.76
5	2008	8.08	34.08
6	2009	4.39	32.11
7	2010	5.88	29.40

The Data from table 2.2 clearly shows that there is a consistent trend of students performing poorly in Biology between the year 2004 and 2010. The percentage of students passing, that is scoring a grade of B+(plus),A-(minus) and A(plain); The grades that indicate a pass in the KCSE is low ranging between 5.88% and 12.03%.

The percentage of students who scored grades D(plain),D-(minus) and E between the year 2004 and 2010; grades that indicate a poor performance in Biology ranged between 29.40% and 49.64%.

The data clearly indicates huge disparity between the percentage of students who perform well and those who perform poorly.

The table below (Table 2.3) obtained from the KNEC 2010 report, shows the percentage of students performance nationally in Biology between the year 2008 and 2010 interms of the grades, the grades range from A which is the best grade to E which is the poorest grade.

Table 2.3. Percentage of students according to grade scored in Biology in the years 2008,2009 and 2010.

SN	GRADE	2008 %	2009 %	2010 %
1.	A	0.27	0.27	0,44
2.	A-	1.71	1.32	1.85
3.	B+	3.1	2.80	3.59
4.	B	4.43	4.48	5.12
5.	B-	6.11	6.54	6.97
6.	C+	8.46	8.86	9.41
7.	C	11.32	11.91	12.35
8.	C-	14.23	14.90	14.79
9.	D+	16.24	16.80	16.01
10.	D	17.78	17.68	16.04
11.	D-	13.96	12.66	11.62
12.	E	2.34	1.77	1.74

The table above (Table 2.3), clearly indicates that there is a consistent increase in the percentage of students who score poor grades in the three years that is 2008, 2009 and 2010.

A study of students performance at KCSE level with special reference to biology was conducted by Lewis (2014). The intention of the study was to investigate students grades in Biology in high schools in Machakos.

The sample comprised of 12 high schools in Machakos, the research used descriptive survey and a stratified sampling method was used to pick the participants. The schools were categorized as either national,county or district schools.
Data was collected from form 4 biology students, biology teachers, heads of science departments and the school head teachers.

The study objectives included:

1. To examine how teachers influence students performance in biology at KCSE level.
2. To determine the influence of socio-economic factors on student grades in biology at KCSE level.

The study results indicated that the performance of students in biology at secondary school level was greatly influenced by syllabus completion, effective practical sessions & the teaching methodology.

Another emerging issue was the long distance that had to be traveled by students in day schools to and from school. These students reached school late and tired thus affecting their concentration in class which had a ripple effect on their performance (Lewis Et al 2014).

2.5.2 PERFORMANCE IN PHYSICS.

Physics is an optional science subject. The subject has continually registered low enrollment in the the national examination that is KCSE

Data in table 1.2 shows that between the years 2010 and 2015, the performance in physics not ignoring the fact that it is still low, has been slightly better than that of other sciences subjects, that is Biology and Chemistry. The national mean grade for physics was 31.5% in the year 2010, 32.94% in 2011, 32.53% in 2012, 36.87% in 2013, 38.84% in 2014 and 43.68% in the year 2015.

In the same year 2015,out of all the students who sat for the physics exam 74,768 scored grades D(plain) and below. This number represents a 12% increase from the year 2014.In the same year (2015) 13,026 students that represents 9% scored grades A(plain) and A-(minus).

The table below (Table 2.4) shows the number of students enrolled to take physics and their percentages compared to the number of students and their percentages enrolled to take the other sciences that is Biology and Chemistry.

Table 2.4. A comparison between the national enrollment to physics and the other science subjects (Biology and Chemistry)between the years 2002 and 2010

YEAR	TOTAL ENTRY	PHYSICS		CHEMISTRY		BIOLOGY	
		ENTRY	%	ENTRY	%	ENTRY	%
2002	197140	54180	27.48	187261	94.99	177328	94.79
2003	206489	56497	27.36	198747	96.25	186403	90.27
2004	220951	60082	27.19	214520	97.09	200797	90.58
2005	259331	69424	27.00	293464	97.74	234975	90.61
2006	252053	72499	29.95	236901	97.87	217928	90.03
2007	270629	83273	30.77	236761	97.29	245911	90.87
2008	300794	92648	30.80	296360	98.50	271735	90.33
2009	338834	104188	30.74	328922	97.07	299304	88.33
2010	357488	109072	30.51	347378	97.17	315063	88.13

source 2002-2011 KCSE examination report

The above table (table 2.4), shows the enrollment of students to all the three science subject (Biology,Chemistry and Physics) nationally, the requirement by the Kenya National Examination Council is that a student takes atleast two science subjects.

The KNEC report shows clearly that most learners are shying away from Physics subject. From the table above it shows that the percentage of students who enrolled for Physics between the years 2002 and 2010 ranges between 27.00% and 30.80% as compared to Chemistry that had an enrollment ranging between 94.99% and 98.5% between the year 2002 and 2010. Biology had an enrollment ranging between 88.13% and 94.79% between the years 2002 and 2010.

Table 2.5. The enrollment to physics interms of gender between the years 2002 and 2010

YEAR	TOTAL ENTRY	FEMALE STUDENTS		MALE STUDENTS	
		ENTRY	%	ENTRY	%
2002	54180	15312	23.26	38868	71.74
2003	56497	16094	28.48	40403	71.52
2004	60082	16966	28.23	43116	71.77
2005	69424	19288	27.78	50136	72.28
2006	72499	21376	29.48	51123	70.52
2007	83273	23767	28.54	59506	71.46
2008	92648	26322	28.41	66326	71.59
2009	104188	29233	28.05	74955	71.95
2010	109072	29964	27.47	79108	72.52

source 2002-2011 KNEC, KCSE examination report

The table above (Table 2.5), summarizes the enrollment to Physics by the two different genders (Female and Male) and it quite clear that most female student drop physics as early as form 1 and 2 as the enrollment for KCSE between the year 2002 and 2010, ranges between 23.26% and 29.48%. Male students dominate in those who take Physics as their enrollment ranges between 70.52% and 71.95% between the year 2002 and 2010.

Studies have shown that fewer students are opting to study Physics beyond form 1 and 2 levels at secondary school (Nderitu,2007). This also trickles down to University level where fewer students are taking physics and its' related courses such as engineering (Lyons,2005). This low enrollment is due to the perception that Physics is a difficult subject (Lyons,2005 and Sperandeo Et al, 2005).

The low enrollment is also attributed to the consistent low performance in the subject (Lyons,2005 and Nderitu,2007). Students low grades in National examinations is likely to bring about low enrollment in classes that follow.

Entry marks, performance at KCPE level in Science will also have an influence on the performance in physics as a science at KCSE level (Orodho,1996). The quality of grades scored in the science paper at KCPE level is highly likely to influence the performance in Science (Biology, Chemistry and physics) subjects as a whole. There is a high correlation between the grades scored in Mathematics and the achievements in Physics(Hudson Et al,2006).

A research conducted Jane by (2011) on the disparities in physics academic achievements and enrollment to physics at high school level.The study objectives included:

1. To identify the disparities in enrollment of students to physics in secondary schools in western Kenya.
2. To identify the disparities in the performance of physics in secondary schools in KCSE.

The sample included secondary schools in western province. The schools were stratified into boys' boarding , girls' boarding and co-education (mixed) schools. Data used was from KCSE 2009 results.
The results indicated that boys performed better than girls in physics also more boys' than girls' take physics as it is an optional subject.

2.5.3 PERFORMANCE IN CHEMISTRY.

Chemistry being another very important science subject that is compulsory in most secondary schools and with the highest enrollment in terms of candidature, according to table 2.3 . The enrollment ranges between 94.99% and 98.50% between the year 2002 and 2010. Despite this fact; Chemistry is the worst performed subject . According to the data in Table 1.1, the performance of Chemistry ranges between 18.995% and 28.11% between the year 2002 and 2010. The subject also recorded same low mean grades. Just like the other science subjects where male students performed better than female students.

Chemistry according to table 1.2 is the poorest performed science subject, in the year 2010 chemistry managed a national mean grade of 22.89%, in 2011 it managed 23.40%,27.72% in 2012, 25.45% in 2013, 32.16% in 2014 and finally 34.36% in 2015.

A research done by Muwanga-zake (1998) showed that the performance in sciences more so chemistry was low due to the following reasons:

1. Teachers misinterpret their own content in the science class.
2. Learners misunderstanding science concepts.
3. There is a barrier between teachers and learners language.

Henerson and Wellington (1998) noted that the greatest barrier to learning Science is Language. Some terminologies that are different in chemistry might mean different in our traditional languages which is not used as a means of teaching and learning Chemistry. In

chemistry smoke, gas and steam are three different things but in the traditional language like 'Kalenjin' they are referred to with one term 'aros'.

A research conducted by Amunga J.K et.al (2009) on disproportion / imbalance in the achievement in chemistry & Biology in high school level.
The fact finding was conducted in 32 secondary schools in western province.

The study had the following objectives:

1. To identify the disproportion in performance in Chemistry & Biology in KCSE between 2005-2009.
2. To explore the factors influencing the distinctive achievements in Chemistry & Biology.

The fact finding mission made use of a descriptive survey design. The data made use of was KCSE 2009 results. ANOVA was used to test the differences between categories (boys' boarding, girls' boarding and mixed schools) performance in chemistry and biology.

2.5.4 PERFORMANCE IN MATHEMATICS.

Mathematics is an important area of learning that is aimed at driving the economies and technological transformation of societies. Mathematics can be used to describe illustrate and interpret numerical patterns of relationship so as to give meaning to various issues in life (National council of curriculum and assessment,2005).

Data from the KNEC 2002 report as indicated in table 1.1, shows that Mathematics performance is even lower than that of the science (Biology, Chemistry and Physics) subjects. The performance ranging between 15.75% and 20.87% between the years 2002 and 2010. Female students registered a poorer performance than their Male counter parts.

Despite the fact the performance in Mathematics has been improving over the years, this performance is still very low. Data in table 1.2 clearly shows that from the year 2010 when mathematics had an overall national mean of 19.17%. this later rose to 21.0% in 2011 and 25.30% in 2012. The performance later dropped to 25.10% in 2013 and 24.02% in 2014. Mathematics registered a slight national improvement to 26.88% in 2015.

Statistics have shown that, nearly 90% of candidates who sat for the mathematics examination scored a mean grade of C-(minus) and below. Only 4% of the candidates managed a mean grade of A-(minus) and A (plain).

Analysis done by the Kenya National Examination Council (KNEC), show that despite the effort by the government the failure rate in Mathematics is still very high and increasing

steadily. For example, in the year 2000, 63.3% of candidates scored grade E, in 2001 the percentage of students scoring grade E rose to 72.2% and in 2002 it rose to 75% (KNEC report,2002).

A study done in Nigeria by James (2014) showed that the factors that contribute to students performance more so in science and mathematics subjects included:

I. Negative attitudes towards the subjects.

II. Lack of resources (both teaching and learning)- these resources include well-equipped laboratories for practicals, Libraries for research and textbooks.

III. Lack of teacher and learner motivation.

IV. Socio-economic back ground.

A study done by Andile & Moses (2006) The study made use of descriptive statistics and non experimental research. It showed that poor performance in secondary school has been attributed to poor teaching methods employed by the teachers, poor motivation of both teachers and the learners, incomplete syllabus coverage, over populated class rooms, the use of English as a mode of instruction which is a second and foreign language to the learners.

A study done by Daniel (2013). The study aimed at establishing if there exists any similarities in the grades obtained in mathematics & chemistry between male & female learners. The objective of the study was to analyze how gender affects the grades obtained by learners in Bomet District in Kenya.

A sample of 208 learners was selected. The study showed that there was a disparity in the grades got in chemistry and mathematics between male and female students where male students out did their female counter parts. This implies that there is a relationship between gender and the grades obtained in mathematics and chemistry Daniel (2013).

2.5.5 PERFORMANCE IN ENGLISH AND SWAHILI.

English and Swahili are both compulsory subjects both at Primary and High school level. Performance in English has been better than that of sciences and mathematics. A comparison with Swahili shows that Swahili is performed better than English.

According to the data on table 1.2, in 2010, English had a national mean grade of 39.26% while Swahili had a mean of 44.34%. In 2011 English had a mean of 36.74% while Swahili had a mean of 49.01. In the year 2012, the mean for English improved to 38.12% while

that of Swahili dropped to 36.32%. The subsequent year had English drop to 35.23% while Swahili improved to 39.91%. In the year 2014, Both English and Swahili improved to 47.68% and 47.58% respectively. In the year 2015, the mean of English dropped to 40.29% while Swahili improved to 47.98%.

English is a second language to most students which is introduced to learners as early as pre-school. A lot of emphasis is placed on learning the language as it is the mode of instruction used in the learning of all the other subjects both at primary and secondary school level.

Claims that poor performance in English language is influenced by the over use of sheng' have been disapproved many organizations, some of such organizations are Uwezo Kenya, Elimu yetu coaltion and also University of Nairobi language department lecturers. Uwezo Kenya through its director Dr.John Mugo, said the over use of sheng' would have affected the performance in Swahili than English.

2.6 SUMMARY OF THE LITERATURE REVIEW

Many researches and studies have been done on the causes of poor performance among learners in both secondary and primary school. Most of the them, concentrated on the school factors, teacher factor and student factors. The main causes of poor performance included the following:

- Quality of teachers / competency and qualification and experience of the teachers.
- Syllabus coverage.
- Location of the school, whether in urban or rural setting.
- Type of school: whether, boys, mixed,girls or even boarding or day.
- Category of the school: whether National, County,Extra county, Sub-county or Private.
- Population of students in the school.
- Teacher student ratio.
- Learning resources.
- Attitude of both the teachers and students.Among a few.

Most of the researchers dwelt on the performance in science subjects only not keeping in mind that languages (Swahili and English) also do contribute to the overall performance of the learners.

Some setbacks to most of the studies done include the following:

1. Most of the studies discussed above made use of very small samples that could not give good inferences for the entire population.
2. Most researches dwelt only on performance in sciences neglecting the performance in other key subjects such as languages.
3. Most of the studies concentrated on the factors affecting the performance and not studied the performance itself.

3 METHODOLOGY

3.1 INTRODUCTION.

This chapter will dwell on the data source, the type of data used, definition of variables to be used in the research, derivation of the model that is the ordinal logistic regression, model assumptions, the sampling frame , sample & sampling technique.

3.2 SOURCE OF DATA.

The data to be used in this research is to be obtained from the Kenya National Examinations council (KNEC) website, some data will be obtained from Kiambu county Education office and the Teachers Service commission (TSC) offices in Kiambu county.

The data to be considered is the KCSE results for the year 2014 and 2015 in Kiambu County.

The data being used is secondary data which is preferred to primary data. Secondary data is data that has been collected previously and can be used by other researchers. The aim of using such data is to increase the sample size and it is also fast to work with.

Some of the common sources of secondary data are: libraries, Government departments, Internet searches and Census reports. One benefit of secondary is that it has already been sorted in an electronic format.

ADVANTAGES OF SECONDARY DATA.

(a.) Secondary sources save time, energy and resources as other people have already collected the data.

(b.) Some times this data has to be purchased, but this price is normally less than the cost of collecting similar data from scratch.

(c.) Longitudinal researches enable the researchers to look at trends and changes in phenomenon over time as this data is frequently collected i.e for my study i can later use data on KCSE from the years 2016,2017 to build up further on the research.

DISADVANTAGES OF SECONDARY DATA.

(a.) It may be strenuous to get data that fits exactly the requirement of the researcher.

(b.) It may be difficult to verify the accuracy of the information.

(c.) The researcher doesn't have control of what the data contains. It could have a lot of irrelevant information to the study being done.

(d.) The data may have not been gotten from a geographical region required for the study, it may also not be from a desired year of study or it may be from a population that is not needed.

(e.) It may be difficult for the researcher to tell the process of data collection & how accurate it was done.

3.3 SAMPLING FRAME.

This is a list of items in a population to be studied. For the case of this study, the sampling frame is the total number of candidates who sat their KCSE in the year 2014 & 2015 country wide and took the subjects: Mathematics, English, Swahili, Biology and Chemistry which are compulsory subjects.

The table below, table 3.1 indicates the number of students who sat for their KCSE in the year 2014 & 2015 distributed in terms of gender and the number of examination centers. Exam centers are areas where the examination was done from. In the year 2014, there were 8646 examination centers. In this centers there were a total of 483630 candidates, out of these, 259746 were male and 223,884 were female candidates. In the year 2015, there were 8057 examination centers that accommodated 522870 candidates, the male candidates were 279289 while the female candidates were 243,581.

Table 3.1. Distribution of Candidates interms of gender in the year 2014 & 2015 Nationally and the number of examination centers.

YEAR	EXAMINATION CENTERS	MALE CANDIDATES	FEMALE CANDIDATES	TOTAL
2014	8646	259746	223884	483630
2015	8057	279289	243581	522870

When we narrow down to Kiambu county that is the area of study. In the year 2014: there were 13,281 female candidates and 13111 male candidates giving a total of 26392, in the

year 2015, there were :14251 female candidates and 13738 female candidates as shown in the table below (table 3.2).

Table 3.2. Distribution of Candidates interms of gender in the year 2014 & 2015 in Kiambu County.

2014			2015		
FEMALE	MALE	TOTAL	FEMALE	MALE	TOTAL
13281	13111	26392	14251	13738	27989

The table below, (table 3.3) shows the number of schools of different categories in Kiambu county in the year 2014 & 2015 from which the sample was obtained. In the year 2014, there were 6 national schools, 30 county schools, 220 sub-county schools and 115 private schools.
For the year 2015, there were 6 national schools, 30 county school, 225 sub-county schools and 110 private schools.

Table 3.3. Number of schools in there different categories in Kiambu county in the year 2014 & 2015..

SCHOOL CATEGORY	2014	2015
NATIONAL	6	6
COUNTY	30	30
SUB-COUNTY	220	225
PRIVATE	115	110
TOTAL	**371**	**371**

Stratified random sampling was made use of to chose the schools and candidates to be used in the study. The strata was based on each of the 12 different sub counties in Kiambu county At the appendix there is a map of Kiambu County showing all the 12 different sub-counties. From each strata schools and candidates were selected randomly. The selection of schools was based on if the school was National, Extra county, county, sub county or private. In both the selected years that is 2014 & 2015, the sample comprised of 23% of the entire population of candidates in the county. The selected schools made up 19% of the schools in the county.

3.4 STATISTICAL SOFTWARE.

The data on KCSE performance comes in Excel.csv format. Excel 2016 program was used to organize the data, select the data of sampled students & selected schools and convert the data into numerical form.
Rstudio/R version 3.4.0 (2017-04-21) and STATA version 13 were both used in the data cleaning and statistical analysis. The software's especially R is readily available and free. With R the output is conveniently stored and can be reviewed later and re-run.
LaTeX.Version Latex2e is the version being used to type this document. The reason for using LaTeX is scientific. LaTeX allows typing of mathematical formula and computations, fitting of tables, aligning of paragraphs and texts.

3.5 DEFINITION OF VARIABLES.

DEPENDENT/ RESPONSE/ OUTCOME/ CRITERION VARIABLE

The dependent variable is the overall performance of students at KCSE level. The grades have been categorized into 5 categories with the range of points for the grades also provided in the table below (table 3.4). The lowest grade which is an 'E' being the reference grade.

Table 3.4. Overall grade categories and range of points

CATEGORY	GRADES	RANGE OF POINTS
1	E	7 – 10
2	D-,D,D+	11 – 31
3	C-,C,C+	32 – 52
4	B-,B,B+	53 – 73
5	A-,A	74 – 84

INDEPENDENT/ PREDICTOR/ INPUT/ EXPLANATORY VARIABLES

Some of the explanatory variables are students grades in 5 subjects that are compulsory and done by most students. The subjects include Mathematics, English, Swahili, Biology and Chemistry.

X_1 – KCSE Performance in Mathematics.
X_2 – KCSE Performance in English.
X_3 – KCSE Performance in Swahili.
X_4 – KCSE Performance in Biology.
X_5 – KCSE Performance in Chemistry.

The subjects are graded A to E, with 12 to 1 point(s) respectively the best grade is an A while the worst grade is an E as described in the table below, (table 3.5).

Table 3.5. Summary of subject grades and points scored for each grade.

GRADES & POINTS DISTRIBUTION			
A	12	C	6
A-	11	C	5
B+	10	D+	4
B	9	D	3
B-	8	D-	2
C+	7	E	1

The other independent variables include: Gender & school type/category.

X_6 – Students gender.

1. Male.
2. Female

X_7 – School type/category.

1. National.
2. Extra County.
3. County.
4. Sub-County.
5. Private.

3.6 ASSUMPTIONS OF THE ORDINAL LOGISTIC REGRESSION MODEL.

The following assumptions are considered when using the ordinal logistic regression/ the proportional odds model:

1. The outcome variable is to be measured at an ordinal level. The ordinal variable has two or several categories that have some natural ordering or ranking.
2. There is one or several input/ explanatory variables that are categorical, Ordinal or Continuous. However, Ordinal input/ explanatory variables must be treated as being either categorical or continuous.
3. There isn't Multicollinearity. Multicollinearity, arises when there are two or several input/ explanatory variables that are very much correlated to each other. This eventually develops an issue with figuring out which variable accounts for the interpretation of the outcome/ response variable. Multicollinearity also brings technical issues when it comes to estimating on ordinal logistic regression.
4. The proportional odds or parallel lines assumption. This means that each Explanatory/ input variable has an indistinguishable (similar) effect on each aggregate (cumulative) split of the ordinal Response/ output variable (Kleinbaum and Klein, 2010).

   Regression coefficients are the same for all categories.
5. Logistic regression never assumes a linear relationship between the output and input variable however a linear relationship is assumed between the logit of the output and input variables.

3.7 ORDINAL LOGISTIC REGRESSION MODEL/ORDINAL REGRESSION MODEL/ PROPORTIONAL ODDS MODEL.

Ordinal Logistic Regression (ordinal Regression) is used to estimate an ordinal output (response) variable given one or several input (explanatory) variable(s). The explanatory variable may be: categorical, interval or a ratio scale variable.

The ordinal regression will enable us to determine which of our independent (predictor) variable(s) have a statistical significant effect on our dependent (response) variable.

In a situation where dependent (explained) categories have an aspect of natural ordering, the model specification must put that into consideration so that the additional information is made use of in the model (Das & Ombui Et al, 2011).

The Proportional Odds model/ Ordinal Logistic Regression Model is a kind of logistic regression where one is modeling the association between input variables and their tendency to be in each higher ordered category.

For example, the model would describe how every input variable individually influences the odds of belonging to the 2^{nd} category or higher 1^{st}; belonging to the 3^{rd} category or higher compared to being in 2^{nd} or 1^{st} category; belonging to the 4^{th} category or higher as opposed to belonging to the 3^{rd},2^{nd} or 1^{st} category ; upto belonging to the 5^{th} category compared to belonging to the 4^{th},3^{rd},2^{nd} or 1^{st} category.

Every comparison has its own cut point/ intercept. However, they share the identical set of regression coefficient. The regression coefficients represent the association of every explanatory (predictor) variable to the odds that an individual would be in each category or above compared to all other lower categories.

In the Ordinal Logistic Regression model, the event being modeled is not having an output in one category as is done in the binary and multinomial models. Rather the event being modeled is having an output in a specific category or any higher category. For instance, for any ordered predicted variable with five categories, the possible events are defined as:

- belonging to category 1,
- belonging to category 2 or 1,
- belonging to category 3, 2 or 1,
- belonging to category 4, 3, 2, or 1,
- belonging to category 5, 4, 3, 2, or 1.

In the Ordinal Logistic Regression model, Every outcome (response) has its own cut point/ intercept but similar regression coefficients. This implies that, the overall odds of an event can differ, but the effect of the regressors on the odds of an event taking place in each subsequent category is similar for every category.

3.7.1 THE MODEL.

We made use of the ordinal logistic regression model to predict the overall performance of students at KCSE level. This is the ordinal dependent variable measured on the 5 category of grades.

$$Y = \begin{cases} 1, & \text{E} \rightarrow \text{Poor.} \\ 2, & \text{D} \rightarrow \text{Weak.} \\ 3, & \text{C} \rightarrow \text{Average.} \\ 4, & \text{B} \rightarrow \text{Good.} \\ 5, & \text{A} \rightarrow \text{Very Good.} \end{cases}$$

Based on independent/predictor variables such as; performance of various subjects such as [Mathematics, English, Swahili, Chemistry & Biology] at KCSE level, students gender [Male, Female], Category of schools [National,Extra County, County, Sub- county & Private].

Here, we will run a set of 4 Binary Logistic Regressions.

* $1 \longrightarrow$ E Versus D,C,B,A.
* $2 \longrightarrow$ E or D Versus C,B,A.
* $3 \longrightarrow$ E,D or C Versus B,A.
* $4 \longrightarrow$ E,D,C or B Versus A.

The proportional odds assumption states that the number added to each set of logarithms to get the next is the same for every case to form an arithmetic sequence.
The reference category is 1 = reference group and so non reference K categories have a linear regression function with regression parameters given as:

$$\beta_j = \beta_{1j}, \beta_{2j}, \beta_{3j}, ..., \beta_{pj} \qquad \text{Where j = 0,1,2,...,k}$$

The set of predictor / independent variable vector with p predictors

$$X_i = (X_{1p}, X_{2p}, X_{3p}, ..., X_{iP}) \qquad \text{Where i = 0,1,2,...,n}$$

In the Ordinal Logistic Regression the dependent variable is an ordered response categorical variable. The independent variable (s) may be categorical, interval or a ratio variable. The response categories have a natural ordering, then the model will be of the form

$$ln\left(\frac{P(Y \leqslant G_j)}{P(Y > G_j)}\right) = \beta_{0j} + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_p X_p \quad \textbf{(3.7.1)}$$

NOTE: If there are m predictors in the response variable then there will be m-1 models which have parallel lines as only the intercept is different, for the study we have 5 categories therefore we will have 4 models.

3.7.2 LINK FUNCTION.

A link function shows the association between the linear predictor and the mean of the distribution. It is a transformation of probabilities that allows for estimation of the model. In ordinal logistic regression we use the Logit link function.

The link function is necessary in a categorical response variable as a categorical response variable is not continuous, is bounded and is not measured on an interval or ratio scale. The link function is the one that differentiates the Logistic regression from the linear regression. It is a function of the mean of the predicted variable Y that is used as the output instead of Y itself.

The link function is the inverse of a distribution function. The Logit link function is the inverse of the standard cumulative logistic distribution function.

$$ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_k X_k \quad \textbf{(3.7.2)}$$

The logit function is the natural log of the odds that Y is equal to one of the categories. The link function is normally a transformation of the probabilities that allow for approximation of the equation. It's main purpose is to connect the random component to the left hand side of the equation to the systematic component on the right hand side of the equation.

1. **Random Component**: This is the probability distribution of the output variable (Y)
2. **Systematic Component**: It specifies the predictor variables $(X_1, X_2, X_3, ..., X_k)$ in the model,moreso their linear combination in creating their linear predictors. e.g $\beta_0 + \beta_1 X_1 + \beta_2 X_2 ....$ in linear regression (Ombui et al,2011).

The logit equation is shown below:

$$log\left[\frac{\pi_1+\pi_2+\pi_3}{\pi_{j+1}+\pi_{j+2}+...+\pi_j}\right]=log\left(\frac{P(Y=j/X_i)}{P(Y=0/X_j)}\right)=\acute{X}_i\beta \tag{3.7.3}$$

from that all the response probabilities are individually determined hence their addition is equal to 1.

$$P(Y=1/X_i),P(Y=2/X_i),...,P(Y=j/X_i)$$
where $j \rightarrow$ the different categories j=1,2,3,4,5
where i= 1,2,3,...,n

$$\sum_{j=1}^{k} P(Y=j/X_i)=1$$

$$\pi_{ij}=\frac{e^{\sum_{k=1}^{j} X_i\beta_k}}{1+e^{\sum_{k=1}^{j} X_i\beta_k}}$$

$$\pi_{ij}=\frac{1}{1+e^{\sum_{k=1}^{j} X_i\beta_k}} \tag{3.7.4}$$

The model will be defined as:

$$ln\left[\frac{P(X_i,X_2,X_3,...,X_p)}{1+P(X_i,X_2,X_3,...,X_p}\right]=\beta_{0j}+\beta_1X_1+\beta_2X_2+...+\beta_pX_p \tag{3.7.5}$$

Which can be re-defined with reference to the individual category output probability by solving (by finding the exponent on both sides) the probabilities to give:

$$P(X_1,X_2,X_3,...,X_p)=\frac{e^{(\beta_{0j}+\beta_1X_1+\beta_2X_2+...+\beta_pX_p)}}{1+e^{(\beta_{0j}+\beta_1X_1+\beta_2X_2+...+\beta_pX_p)}} \tag{3.7.6}$$

For any given regressor X_i, the coefficient β_j shows the change in the log odds of the regressed relationship with each unit change(increase) in the variable X_i. In random fixed values for the remaining regressors $X_1,X_2,X_3,...,X_p$. The exponential regression coefficient exponential β_j gives the odds ratio that is associated with each unit increase/decrease in X_i (O'connel A.A, 2006).

The slope coefficients (β's) are interpreted as the effect of a unit increase / decrease in the x (input) variable on the outcome logit with all the other variables in the model/equation held constant or adjusted for.

3.7.3 THE ODDS RATIO IN ORDINAL LOGISTIC REGRESSION.

The odds ratio in ordinal logistic regression is given by the function:

$$\pi_{ij} = \frac{e^{\sum_{k=1}^{j} X_i \beta_k}}{1 + e^{\sum_{k=1}^{j} X_i \beta_k}} \qquad where\, j = 1,2,3,4,5$$

$$1 - \pi_{ij} = \frac{1}{1 + e^{\sum_{k=1}^{j} X_i \beta_k}} \tag{3.7.7}$$

To simplify the notation, the individual probabilities for π_{ij}, for the odds to be given as:
Let Odds Ratio (OR) be η

$$\eta_j = \frac{\pi_{ij}}{1 - \pi_{ij}} = \frac{\frac{e^{\sum_{k=1}^{j} X_i \beta_k}}{1 + e^{\sum_{k=1}^{j} X_i \beta_k}}}{\frac{1}{1 + e^{\sum_{k=1}^{j} X_i \beta_k}}}$$

$$\Rightarrow \frac{\pi_{ij}}{1 - \pi_{ij}} = e^{X\beta} \tag{3.7.8}$$

By computing or obtaining the natural logarithm on each side of equation 3.7.8 above we get.

$$ln\left[\frac{\pi_{ij}}{1 - \pi_{ij}}\right] = X\beta$$

where:

$$X\beta = \beta_0 + X_1\beta_1 + X_2\beta_2 + ... + X_i\beta_j$$

$$\therefore \qquad ln\left[\frac{\pi_{ij}}{1 - \pi_{ij}}\right] = \beta_0 + X_1\beta_1 + X_2\beta_2 + ... + X_i\beta_j \tag{3.7.9}$$

From that we obtain the derivative to get the relationship between the estimated parameters and the odds ratio:

$$\frac{\mathrm{d}}{\mathrm{d}X} ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_j \tag{3.7.10}$$

The parameter β_j shows the percentage increase/decrease in the log-odds ratio from each unit increase/decrease in one of the predictor variables X_i. To establish the relationship between the odds ratios and the probability of π, we have:

$$log-odds = \ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \ln\left(\pi_{ij}\right) - \ln\left(1-\pi_{ij}\right) \tag{3.7.11}$$

on differentiating you obtain:

$$\frac{\mathrm{d}(log-odds)}{\mathrm{d}X} = \frac{1}{\pi_{ij}} - \frac{1}{(1-\pi_{ij})^2}$$

where,

$$\frac{1}{\pi_{ij}} - \frac{1}{(1-\pi_{ij})^2} = \ln\left(\frac{\pi_{1j}}{1-\pi_{1j}}\right) - \ln\left(\frac{\pi_{0j}}{1-\pi_{0j}}\right) = \beta_{ij} \tag{3.7.12}$$

From the above,π_{ij} and π_{0j} are the probabilities in defining the change in the log-odds ratio.

$$\therefore \qquad \ln\left\{\frac{\frac{\pi_{1j}}{1-\pi_{1j}}}{\frac{\pi_{0j}}{1-\pi_{0j}}}\right\} = \beta_j$$

By introducing the natural logarithm (ln) on both sides of equation above, we obtain:

$$\left\{\frac{\frac{\pi_{1j}}{1-\pi_{1j}}}{\frac{\pi_{0j}}{1-\pi_{0j}}}\right\} = e^{\beta_j} \tag{3.7.13}$$

The equation above represents the odds-ratio. It is given in ratio form of the individual probability odds ratio of the non reference categories of the response variable, compared to the reference category of the response variable.

Odds Ratio (η_j) are used to interpret model parameters. If the final grades can be ranked as E(1),D(2),C(3),B(4) & A(5), the following odds can be modeled.

$$\theta_1 = \frac{prob(grade\ E)}{prob(grade\ D\ or\ Cor\ B\ or\ A)}$$

$$\theta_2 = \frac{prob(grade\ E\ or\ D)}{prob(grade\ C\ or\ B\ or\ A)}$$

$$\theta_3 = \frac{prob(grade\ E\ or\ D\ or\ C)}{prob(grade\ B\ or\ A)} \tag{3.7.14}$$

$$\theta_4 = \frac{prob(grade\ E\ or\ D\ or\ C\ or\ B)}{prob(grade\ A)}$$

The last category (A) does not have an odds associated with it since the probability of scoring upto and including the last score is 1. Hence, this category serves as the base or reference category.

3.7.4 THE GOODNESS OF FIT TEST.

This test is important as it helps to tell if the data that has been sampled fits a given distribution. It normally measures how a random sample matches with a theoretical probability distribution function. It shows if a data that has been sampled shows what is expected to be found in the actual population. It measures how well the observed data corresponds to the fitted model.

The general procedure is to define the test statistic that is a function for the data. We then find the distance between the hypothesis & the data and then evaluate the probability of getting data that has a large value of the specified test statistic compared to the value observed. If this hypothesis is true, this probability is referred to as the confidence level. Commonly used goodness of fit tests include:

(a.) Pearson's chi-square (χ^2) test.

(b.) Hosmer - Lemeshow (C_g) test.

(c.) Deviance Statistics.

(d.) Lipsitz likelihood ratio test.

For the purpose of this project i will just mention a few.

3.7.4.1 PEARSONS CHI-SQUARE (χ^2) TEST.

This is a goodness of fit test that is used by analysts who desire to tell how well the model fits the data. It tests/shows the disparity between the reduced model and the full model. It shows how “close" the observed values are to those which are expected in the fitted model. This test is used to tell if the predicted values deviate from the observed values in a way that the distribution does not predict. The χ^2 test, tests the association of the variables in the two-way table, where the assumed model of independence is evaluated against the observed data. The general form of the χ^2 test is;

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected} \tag{3.7.15}$$

Incase the obtained test statistic is large, this implies that observed and expected values are not close and the model is a poor fit to the data.
If $\chi^2_{\ calculated} \geqslant \chi^2_{\ tabulated}$ you reject the null hypothesis (H_0) at the given α level of significance.

However, the test isn't important if the number of distinct values is approximately equal to the number of observations but is important if you have multiple observations at the same values of the predictors.

If the *p*-value for the test is below the chosen significance level, the predicted probabilities will deviate from the observed probabilities in a way that the distribution can not predict. Some of the common reasons for this kind of a deviation could be:

. A Link function that is not correct.

. The omission higher-order terms for variance in that model.

. Omission of predictors from the model.

3.8 ESTIMATION OF THE MODEL PARAMETERS.

There are several methods that are used to estimate model parameters ; These methods include:

1. Maximum likelihood estimation (MLE). MLE is applied when you want to maximise the total probability of the data.
2. Optimizing some criterion of fit (e.g minimizing the χ^2 , methods of least squares (in the method least squares, we want to find the line that minimizes the total squared distance between the data points and the regression line) and minimizing the goodness of fit statistic.
3. Method of moment (Moment matching).
4. Quartile matching.
5. Bayensian methods (MAP - maximum a posterior estimation and MMSE - minimum mean square error)

The commonly used method for estimating the parameters is the Maximum Likelihood Estimator. This is a method of determining the values of parameters in a model. The value of parameters are found such that they maximize the likelihood that the process described by the model produce, the data that was observed.

The fitted model will be assessed using the wald χ^2 statistic and the maximum likelihood ratio test. (Shakhawt et.al ,2012)

The likelihood function of estimating $\underset{\sim}{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_k)'$ is given by:

$$L(\underset{\sim}{\beta}) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

where

$$p = \frac{exp(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij})}{1 + exp(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij})}$$

Therefore replacing the equation above with the value of p we obtain:

$$L(\underset{\sim}{\beta}) = \prod_{i=1}^{n} \left(\frac{exp(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij})}{1 + exp(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij})} \right)^{y_i} \left(\frac{1}{1 + exp(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij})} \right)^{1-y_i}$$

In the maximization process, we have to obtain the log of the likelihood function $L(\underset{\sim}{\beta})$

The best thing about the maximizing of the log likelihood function $L(\underset{\sim}{\beta})$ instead of the likelihood is that the log likelihood function is the sum function while the likelihood function is a product function.

Therefore, the log likelihood function will be given by:

$$\ell(\underset{\sim}{\beta}) = \sum_{i=1}^{n}(\beta_0 + \sum_{j=1}^{k}\beta_j X_{ij})y_i - \sum_{i=1}^{n} log(1 + exp(\beta_0 + \sum_{j=1}^{k}\beta_j X_{ij}))$$

The maximum likelihood equation for obtaining the $\underset{\sim}{\beta}$ will be obtained differentiating with respect to β_0 and β_j then equating to zero as shown below.

$$\frac{\partial \ell(\underset{\sim}{\beta})}{\partial \underset{\sim}{\beta_0}} = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{exp(\beta_0 + \sum_{j=1}^{k}\beta_j X_{ij})}{1 + exp(\beta_0 + \sum_{j=1}^{k}\beta_j X_{ij})} = 0$$

$$\frac{\partial \ell(\underset{\sim}{\beta})}{\partial \underset{\sim}{\beta_j}} = \sum_{i=1}^{n} y_i x_{ij} - \sum_{i=1}^{n} x_{ij} \frac{exp(\beta_0 + \sum_{j=1}^{k}\beta_j X_{ij})}{1 + exp(\beta_0 + \sum_{j=1}^{k}\beta_j X_{ij})} = 0 \qquad \textbf{(3.8.1)}$$

where j= 1,2,3,...,k

Due to the fact that the equations are non-linear, they require an iterative solution which is readily available in the statistical software.

The entire process of maximizing and obtaining the estimated coefficients is concluded after obtaining the 1st and 2nd derivatives of the log-likelihood function and ensuring that the derivative are equal to zero for a maximum.

3.9 ASSESSING THE FITTED MODEL.

This entails assessing the appropriateness or usefulness of the fitted model. It is normally done after estimating the regression coefficients. In this process, we first test the importance of each explanatory variables. This assessment is done using the wald χ^2 statistic or the likelihood ratio test.
After testing the importance of each explanatory variables, you test the overall goodness of fit of the model.

3.9.1 DEVIANCE STATISTIC OR -2 LOG-LIKELIHOOD STATISTIC.

The deviance is a measure of how much variation that is not explained is in a logistic regression model. The larger the value, the less precise the model is.
Deviance statistic shows comparison between the difference in probability, the predicted outcomes and actual outcome for each case and cumulative sums of these differences to provide an estimate of the total error in the model.It indicates of how much information information that is not explained is in the model after it has been fitted. A large value shows a model that has been poorly fitted.

Deviance statistic can also he used in a model to check if adding or removing certain predictor variables will change its predictive power.

If the deviance statistic is reducing to a statistically significant magnitude with each set of input variables added to the model, then it is improving at precisely predicting the outcome of each case. The formula for deviance statistic is given below:

$$\begin{aligned} Deviance\ \ statistic,\ \ D &= -2ln\left[\frac{L_c}{L_s}\right] \\ &= (-2)ln(L_c) - (-2)ln(L_s) \end{aligned} \tag{3.9.1}$$

where $L_c \Rightarrow$ is the likelihood of data using current model,
$L_s \Rightarrow$ is the likelihood of data using standard model.

From the above formula, the larger the deviance,the poorer the fit to the model and the smaller the deviance the better the fit to the model.

3.10 THE MEASURE OF PREDICTIVE POWER.

The main method for determining the predictive power of a model is the use of R-squared (R^2). There are three types of (R^2), and that is ;

1. Mc Fadden (R^2) or Pseudo (R^2).
2. Cox - snell (R^2).
3. Nagelkerke's (R^2).

3.10.1 Mc FADDEN (R^2)

.

$$R^2_{McF} = 1 - ln(L_m)/ln(L_o) \tag{3.10.1}$$

where $ln \Rightarrow$ is defined as the natural logarithm.
In the equation, $ln(L_o)$ has an analogous role to the residual sum of squares in linear regression. Thus, the equation matches to the proportional reduction in the error variance (Mc Fadden, 1974).

3.10.2 THE COX AND SNELLS (R^2)

.

$$R^2_{c\&s} = 1 - \left(l_m/l_0\right)^{2/n} \tag{3.10.2}$$

where $n \Rightarrow$ is the size of the sample.
This equation is to be used in normal theory linear regression. It is an identity implying, that the usual (R^2) for linear regression relies on the likelihood for the models with or without explanatory/ input variables. (Cox et.al, 1989).

3.10.3 NAGELKERKE (R^2)

Nagelkerke (R^2) is normally computed by dividing cox and snell (R^2) by its maximum value.

$$Nagelkerke \quad (R^2) = \frac{1 - \left[\frac{L(m_{intercept})}{L(m_{full})}\right]^{2/n}}{1 - L(m_{intercept})^{2/n}} \tag{3.10.3}$$

4 DATA ANALYSIS AND RESULTS

4.1 THE MODEL.

$$ln\left(\frac{P(Y\leqslant G_j)}{P(Y>G_j)}\right)=\beta_{0j}+\beta_1X_1+\beta_2X_2+\beta_3X_3+...+\beta_pX_p \quad \textbf{(4.1.1)}$$

MODEL 1 FOR THE 2014 KCSE RESULTS.

$$ln\left(\frac{P(Y=E)}{P(Y=D\,or\,C\,or\,B\,or\,A)}\right)=61.4793+0.9997\,MATHS+1.1607\,ENG+1.2880\,KIS+1.3332\,BIO$$
$$+1.1803\,CHEM+0.1270\,FEMALE-0.2565\,EXTRA\,COUNTY-0.5683\,COUNTY$$
$$-0.8559\,SUB\,COUNTY-1.2047\,PRIVATE. \quad \textbf{(4.1.2)}$$

MODEL 2 FOR THE 2014 KCSE RESULTS.

$$ln\left(\frac{P(Y=D\,or\,E)}{P(Y=C\,or\,B\,or\,A)}\right)=41.4124+0.9997MATHS+1.1607ENG+1.2880KIS+1.3332BIO$$
$$+1.1803CHEM+0.1270FEMALE-0.2565EXTRACOUNTY-0.5683COUNTY$$
$$-0.8559SUBCOUNTY-1.2047PRIVATE. \quad \textbf{(4.1.3)}$$

MODEL 3 FOR THE 2014 KCSE RESULTS.

$$ln\left(\frac{P(Y=C\,or\,D\,or\,E)}{P(Y=B\,or\,A)}\right)=23.0308+0.9997MATHS+1.1607ENG+1.2880KIS+1.3332BIO$$
$$+1.1803CHEM+0.1270FEMALE-0.2565EXTRACOUNTY-0.5683COUNTY$$
$$-0.8559SUBCOUNTY-1.2047PRIVATE. \quad \textbf{(4.1.4)}$$

MODEL 4 FOR THE 2014 KCSE RESULTS.

$$ln\left(\frac{P(Y=B\,or\,C\,or\,D\,or\,E)}{P(Y=A)}\right) = 6.7920 + 0.9997MATHS + 1.1607ENG + 1.2880KIS + 1.3332BIO + 1.1803CHEM + 0.1270FEMALE - 0.2565EXTRACOUNTY - 0.5683COUNTY - 0.8559SUBCOUNTY - 1.2047PRIVATE. \quad (4.1.5)$$

4.1.1 Overall Significance of the model.

The model was found to be significant overaly since the P-Value of the entire model was 0.0000 which is less than 0.05 (alpha). Meaning that the model does fit the data well

Table 4.1. Study Variable Maximum Likelihood Parameter Estimates for 2014

PARAMETER	ESTIMATES	ODDS RATIO	STD ERROR	P-VALUE	t-value	CONFIDENCE INTERVAL
intercept a	6.7920		0.4404	1.1726×10^{-53}	15.4215	LOWER / UPPER
intercept b	23.0308		0.8132	1.8496×10^{-176}	28.32215	
intercept c	41.4124		1.3406	1.6344×10^{-209}	30.8899	
intercept d	61.4793		1.9331	5.917×10^{-222}	31.8028	
MATHS	0.9997	2.7174	0.0426	4.7743×10^{-176}	23.493	(2.5048,2.9597)
ENGLISH	1.1608	3.1923	0.0611	1.8343×10^{-80}	18.996	(2.8380,3.6065)
SWAHILI	1.2880	3.6256	0.0544	8.6742×10^{-124}	23.663	(3.2660,4.3046)
BIOLOGY	1.3332	3.7933	0.0632	8.641×10^{-99}	21.096	(3.3597,4.3046)
CHEMISTRY	1.1803	3.2552	0.0544	7.645×10^{-95}	20.662	(2.5048,2.9597)
GENDER.FEMALE	0.1270	1.1354	0.0.1246	3.081×10^{-1}	1.019	(0.8894,1.4498)
SCHL CATE E COUNTY	-0.2565	0.7738	0.1775	1.4833×10^{-1}	-1.445	(0.5459,1.0949)
COUNTY	-0.5683	0.5664	0.2230	1.0807×10^{-2}	-2.549	(0.3655,0.8762)
SUB COUNTY	-0.8559	0.4249	0.2311	2.1268×10^{-4}	-3.703	(0.2697,0.6676)
PRIVATE	-1.2047	0.2998	0.3337	1.1726×10^{-53}	-3.610	(0.1554,0.5745)

RESIDUAL DEVIANCE = 2088.539 AIC = 2116.539

4.1.2 INTERPRETATION.

From the results obtained in table 4.1 above we will not report on Gender and School category that is Extra county schools because they have a P-Value greater than 0.05. The two are not significant predictors, their P-values are 0.3081 and 0.1483 respectively.

MATHEMATICS.
For every increase in grade in Mathematics a student is 2.7174 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

ENGLISH.
For every increase in grade in English a student is 3.1923 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

SWAHILI.
For every increase in grade in Swahili a student is 3.6256 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

BIOLOGY.
For every increase in grade in Biology a student is 3.7933 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

CHEMISTRY.
For every increase in grade in Chemistry a student is 3.2552 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

SCHOOL CATEGORY.

COUNTY SCHOOLS.
A student who attends a county school is $(1-0.5664)\times 100\% = 43.36\%$ less likely to perform better than a student who attends a national school. Adjusting for all the other predictors (holding all the other predictors constant).

SUB COUNTY SCHOOLS.
A student who attends a sub county school is $(1-0.4249)\times 100\% = 57.51\%$ less likely to perform better than a student who attends a national school. Adjusting for all the other predictors (holding all the other predictors constant).

PRIVATE SCHOOLS.
A student who attends a Private school is $(1-0.2998)\times 100\% = 70.02\%$ less likely to perform better than a student who attends a national school. Adjusting for all the other predictors (holding all the other predictors constant).

Table 4.2. Subjects that contribute the most to students overall performance in order of merit 2014

SUBJECT	ODDS RATIO
Biology	3.7933
Swahili	3.6256
Chemistry	3.2552
English	3.1923
Mathematics	2.7174

Table 4.2 shows the subjects in order of merit that contribute the most to students overall mean grade. Biology tops the list with an odds ratio of 3.7933 followed by Swahili with an odds ratio of 3.6256. Chemistry follows with an odds ratio of 3.2552 followed by English and finally Mathematics that have odds ratios of 3.1923 and 2.7174 respectively.

4.1.3 Model Fitting Information .

Table 4.3. Year 2014 Model Fiting Information.

Model	-2 log likelihood	Chi-square	d.f	sig
Intercept only	16754.075			
Final	0.0000	16754.075	10	0.0000

The P-Value of 0.000 which is greater than 0.05, in table 4.3 above shows that the model is overally significant.

Table 4.4. Year 2014 Model Goodness-of-fit.

	Chi-square	d.f	sig
Pearson	2317.640	20002	1.000
Deviance	1947.680	20002	1.000

The hypothesis being tested is as follows:
H_0: The data that is observed is consistent with the model that has been fitted. V/S
H_1: The data that is observed is not consistent with the with the model that has been fitted.
The table above (table 4.4) on the goodness on fit test shows that the data fits the model well since we have a p-value of 1.0 that is more than 0.05 we fail to reject the null hypothesis (H_0) at 5% level of significance and conclude that the data that has been observed is consistent with the estimated values in the ordinary logistic regression.

Table 4.5. Year 2014 Psedo R-Squared.

Cox & Snell	0.931
Nagelkerke	0.991
Mc Fadden	0.993

R^2 Explains the percentage change in variation of the response variable that has been explained by the explanatory/input variables.

For the table above (table 4.5), More than 90% of the change in variation in the response/ outcome variable is explained by the explanatory/ input variables or is explained by the model.

MODEL 1 FOR THE 2015 KCSE RESULTS.

$$ln\left(\frac{P(Y=E)}{P(Y=D\,or\,C\,or\,B\,or\,A)}\right)=56.0823+1.0487\,MATHS+1.0865\,ENG+1.2126\,KIS+1.0102\,BIO$$
$$+1.0948\,CHEM+0.7659\,FEMALE-0.3283\,EXTRA\,COUNTY-0.8788\,COUNTY$$
$$-1.2608\,SUB\,COUNTY-1.3332\,PRIVATE. \quad \textbf{(4.1.6)}$$

MODEL 2 FOR THE 2015 KCSE RESULTS.

$$ln\left(\frac{P(Y=D\,or\,E)}{P(Y=C\,or\,B\,or\,A)}\right)=38.0063+1.0487\,MATHS+1.0865\,ENG+1.2126\,KIS+1.0102\,BIO$$
$$+1.0948\,CHEM+0.7659\,FEMALE-0.3283\,EXTRA\,COUNTY$$
$$-0.8788\,COUNTY-1.2608\,SUB\,COUNTY-1.3332\,PRIVATE. \quad \textbf{(4.1.7)}$$

MODEL 3 FOR THE 2015 KCSE RESULTS.

$$ln\left(\frac{P(Y=C\,or\,D\,or\,E)}{P(Y=B\,or\,A)}\right)=20.8328+1.0487\,MATHS+1.0865\,ENG+1.2126\,KIS+1.0102\,BIO$$
$$+1.0948\,CHEM+0.7659\,FEMALE-0.3283\,EXTRA\,COUNTY-0.8788\,COUNTY$$
$$-1.2608\,SUB\,COUNTY-1.3332\,PRIVATE. \quad \textbf{(4.1.8)}$$

MODEL 4 FOR THE 2015 KCSE RESULTS.

$$ln\left(\frac{P(Y=B\,or\,C\,or\,D\,or\,E)}{P(Y=A)}\right)=5.1739+1.0487\,MATHS+1.0865\,ENG+1.2126\,KIS+1.0102\,BIO$$
$$+1.0948\,CHEM+0.7659\,FEMALE-0.3283\,EXTRA\,COUNTY-0.8788\,COUNTY$$
$$-1.2608\,SUB\,COUNTY-1.3332\,PRIVATE. \quad \textbf{(4.1.9)}$$

4.1.4 Overall Significance of the model.

The model was found to be significant overaly since the P-Value of the entire model was 0.0000 which is less than 0.05 (alpha).

This implies that the model fits the data well

Table 4.6. Study Variable Maximum Likelihood Parameter Estimates For 2015

PARAMETER	ESTIMATES	ODDS RATIO	STD ERROR	P-VALUE	t-value	CONFIDENCE INTERVAL
intercept a	5.1739		0.4278	1.1×10^{-33}	12.093	LOWER / UPPER
intercept b	20.8328		0.7326	6.8×10^{-178}	28.438	
intercept c	38.0063		1.2126	1.2×10^{-215}	31.343	
intercept d	56.0823		1.7301	1.7×10^{-230}	32.415	
MATHS	1.0487	2.8539	0.0438	2.1×10^{-126}	23.916	(2.6237,3.1160)
ENGLISH	1.0865	2.9640	0.0571	1.1×10^{-80}	19.025	(2.6551,3.3216)
SWAHILI	1.2126	3.3624	0.0510	3.9×10^{-125}	23.793	(3.0491,3.7237)
BIOLOGY	1.0102	2.7461	0.0493	2.23×10^{-93}	20.498	(2.4974,3.0299)
CHEMISTRY	1.0948	2.9886	0.0504	1.4×10^{-104}	21.717	(2.7124,3.3083)
GENDER.FEMALE	0.7659	2.1510	0.1200	1.73×10^{-10}	6.384	(1.7024,2.7253)
SCHL CATE E COUNTY	-0.3283	0.7202	0.1690	5.2×10^{-2}	-1.943	(0.5168,1.0026)
COUNTY	-0.8788	0.4153	0.2190	5.9×10^{-5}	-4.014	(0.2700,0.6372)
SUB COUNTY	-1.2608	0.2834	0.2350	8.0×10^{-8}	-5.366	(0.1785,0.4486)
PRIVATE	-1.3332	0.2636	0.3162	2.5×10^{-5}	-4.216	(0.1417,0.4897)

RESIDUAL DEVIANCE = 2294.214 AIC = 2322.214

4.1.5 INTERPRETATION.

The interpretation of analysis data obtained in table 4.6 is as follows;

MATHEMATICS.
For every increase in grade in Mathematics a student is 2.8539 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

ENGLISH.
For every increase in grade in English a student is 2.9640 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

SWAHILI.
For every increase in grade in Swahili a student is 3.3624 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

BIOLOGY.
For every increase in grade in Biology a student is 2.7461 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

CHEMISTRY.
For every increase in grade in Chemistry a student is 2.9886 times more likely to improve on their overall grade. Adjusting for all the other predictors (holding all the other predictors constant).

GENDER.
Female students are 2.1510 times more likely to perform better than their male counter parts. Adjusting for all the other predictors (holding all the other predictors constant).

SCHOOL CATEGORY.
We will not report for Extra county schools since it has a P-Value of 0.052 which is slightly above the α (0.05) level of significance.

COUNTY SCHOOLS.
A student who attends a county school is $(1-0.4148)\times 100\% = 58.47\%$ less likely to perform better than a student who attends a national school. Adjusting for all the other predictors (holding all the other predictors constant).

SUB COUNTY SCHOOLS.
A student who attends a sub county school is $(1-0.2834)\times 100\% = 71.66\%$ less likely to perform better than a student who attends a national school. Adjusting for all the other predictors (holding all the other predictors constant).

PRIVATE SCHOOLS.
A student who attends a Private school is $(1-0.2636)\times 100\% = 73.64\%$ less likely to perform better than a student who attends a national school. Adjusting for all the other predictors (holding all the other predictors constant).

Table 4.7. Subjects that contribute the most to students overall performance in order of merit 2015

SUBJECT	ODDS RATIO
Swahili	3.3624
Chemistry	2.9886
English	2.9640
Mathematics	2.8539
Biology	2.7461

Table 4.7 above shows the subjects in order of merit that contribute the most to students overall mean grade. Kiswahili tops the list with an odds ratio of 3.3624 followed by Chemistry with an odds ratio of 2.9886. English follows with an odds ratio of 2.9640 followed by Mathematics and finally Biology that have odds ratios of 2.8539 and 2.7461 respectively.

4.1.6 Model Fitting Information .

Table 4.8. Year 2015 Model Fiting Information.

Model	-2 log likelihood	Chi-square	d.f	sig
Intercept only	16438.347			
Final	0.0000	16438.347	10	0.0000

The P-Value of 0.000 which is bigger than 0.05, in table 4.8 above shows that the model is overally significant.

Table 4.9. Year 2015 Model Goodness-of-fit.

	Chi-square	d.f	sig
Pearson	2179.649	19670	1.000
Deviance	2103.165	19670	1.000

The hypothesis being tested are as follows:
H_0: The data that is observed is consistent with the model that has been fitted.
H_1: The data that is observed is not consistent with the with the model that has been fitted.
The table above (table 4.9) on the goodness on fit test shows that the data fits the model well since we have a p-value of 1.0 that is more than 0.05 we fail to reject the null hypothesis H_0 at 5% level of significance and conclude that the data that has been observed is consistent with the estimated values in the ordinary logistic regression.

Table 4.10. Year 2015 Psedo R-Squared.

Cox & Snell	0.931
Nagelkerke	0.998
Mc Fadden	0.992

R^2 Explains the percentage change in variation in the response variable that is explained by the explanatory/input variables.

For the table above (table 4.10), More than 90% of the change in variation in the response/outcome variable is explained by the explanatory/input variables or is explained by the model.

4.1.7 SUMMARY OF STUDENTS OVERALL AND SUBJECT PERFORMANCE WITH REFERENCE TO GENDER FOR THE YEAR 2014.

Table 4.11. Overall mean grades.

GRADE / GENDER	*A*	*B*	*C*	*D*	*E*	*TOTAL*
MALE	495	851	966	621	25	2958
FEMALE	451	1307	1136	409	11	3314
TOTAL	946	2158	2122	1030	36	6272

χ^2 = 141.48 df = 4 P-Value < 2.2×10^{-16}

From the table above, (table 4.11): More male students scored grade A, D & E than female students. The number of female students scoring grade B & c was higher than that of male students. Approximately 15% of the students in the sample scored grade A, 34% scored B & C, 16% scored grade D and 1% scored grade E.

Table 4.12. Mathematics performance.

GRADE / GENDER	*A*	*A*−	*B*+	*B*	*B*−	*C*+	*C*	*C*−	*D*+	*D*	*D*−	*E*
MALE	510	161	160	134	146	179	145	151	176	328	402	466
FEMALE	574	198	210	204	201	242	227	244	165	353	339	357

χ^2 = 88.105 df = 11 P-Value < 3.97×10^{-14}

From the table above, (table 4.12): More female than male students in the sample scored grade c+ and above while more male students than female scored grade c and below. 17% of the students scored grade A, 6% A-, B+, B-,& C respectively, 7% scored C+, C- respectively, 5% scored D+, 11% scored D, 12% scored D- and 13% scored grade E.

Table 4.13. English performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	97	185	436	503	510	406	242	223	107	186	54	9
FEMALE	190	341	792	747	515	287	110	113	49	134	34	2

χ^2 = 353.15 df = 11 P-Value < 2.2×10^{-16}

The table above (table 4.13) shows that from the sample used, more female than male students scored C+ and above while more male than female students scored C and below in English. The percentage of students scoring the different grades was as follows; 5% of all the students scored grade A, C- were 5% D, 8% A-, 20% B+ & B respectively, 16% B-, 11% C+, 6% C, 2% D+ , 1% D- and those who scored E were less than 1%.

Table 4.14. Swahili performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	293	217	314	249	263	437	299	257	184	315	109	21
FEMALE	234	256	522	415	407	543	336	224	124	192	57	4

χ^2 = 199.72 df = 11 P-Value < 2.2×10^{-16}

The table above (table 4.14) with the performance of Swahili shows that; More female than male students scored grade C+ and above while more male than female students scored C and below. The percentage of students with respect to grades was as follows: 8% scored A, A-, C- & D respectively, 13% B+, 11% B& B- respectively, 16% C+, 10% C, 5% D+ and D- respectively & E both had less than 3%.

Table 4.15. Biology performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	412	162	180	211	284	246	274	244	169	399	274	103
FEMALE	156	162	268	341	363	389	407	370	221	361	197	79

χ^2 = 262.19 df = 11 P-Value < 2.2×10^{-16}

The table above (Table 4.15) on the performance in Biology, the percentage of students who scored the different grades from the sample used was as follows:9% scored grades A and B respectively, 5% A-, 7% B+, 10% B-, C+ & C- respectively, 11% C, 6% D+, 12% D , 8% D- and 3% E.

Table 4.16. Chemistry performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	325	146	168	208	172	257	239	236	248	547	329	83
FEMALE	268	149	202	232	195	371	282	318	357	599	287	54

χ^2 = 58.747 df = 11 P-Value < 1.583×10^{-8}

Table 4.16 on the performance in chemistry shows that the percentage of students who scored the different grades was as follows: 9% A & C- respectively, 5% A-, 6% B+ & B- respectively, 7% B, 10% C+, D+ & D- respectively, 8% C, 18% D and 2% E.

Table 4.17. Summary of school category and number of students with reference to gender.

SCHOOL CATEGORY	FEMALE	MALE	TOTAL
National	635	831	1466
Extra County	587	1223	1810
County	621	608	1229
Sub County	993	522	1515
Private	122	130	252
TOTAL	2958	3314	6272

χ^2 = 377.51 df = 4 P-Value < 2.2×10^{-16}

Table 4.17 shows that 23% of the sample came from National schools out of which 43% were female students while 57% were male students, 29% were from Extra county schools of which 48% were female students and 52% male students. County schools took up 20% of the entire sample of which female students were 51% and 49% were male students. Sub county schools contributed 24% of the sample of which 66% were female students and 36% male students. Private schools took up 4% of the sample of which 48% were female students and 52% male students.

Table 4.18. Summary of school category and overall grades attained.

SCHOOL CATEGORY	A	B	C	D	E	TOTAL
National	717	618	106	25	0	1466
Extra County	195	907	672	36	0	1810
County	13	314	726	176	0	1229
Sub County	8	242	518	731	16	1515
Private	13	77	80	62	20	252
TOTAL	946	2158	2102	1030	36	6272

χ^2 = 4048.3 df = 16 P-Value < 2.2×10^{-16}

Table 4.18 with a summary of school category and overall grades scored by students shows that : 96% of students who got a mean grade of A were from National and Extra county schools,while 4% were from county, sub-county and private schools. 71% of those who scored grade B were from National and Extra county schools, while 29% from county, sub-county and private schools. 37% from both National and Extra county schools scored grade C, while 63% of those who scored C were from county, sub-county and private schools. 6% of students from national and Extra county schools managed to score grade

D, while 94% were from county, sub-county and private schools. All the students who scored a mean grade of E were from sub-county and private schools.

4.1.8 SUMMARY OF STUDENTS OVERALL AND SUBJECT PERFORMANCE WITH REFERENCE TO GENDER FOR THE YEAR 2015.

Table 4.19. Overall mean grades.

GRADE / GENDER	*A*	*B*	*C*	*D*	*E*	*TOTAL*
MALE	562	836	886	594	13	2891
FEMALE	445	1228	1094	470	7	3244
TOTAL	1007	2064	1980	1064	20	6135

χ^2 = 106.19 df = 4 P-Value < 2.2×10^{-16}

Table 4.19 above shows that more male than female students scored grade A,D and E More female than male students scored grade B and C.56% of those who scored A were males while 44% were female. 41% of those who scored B were male while 59% were female. 45% of those who scored C were male while 55% were female. 56% of those who scored D were male while 44% were female. 65% of those who scored a mean grade of E were male while 35% were female.

Table 4.20. Mathematics performance.

GRADE / GENDER	*A*	*A*−	*B*+	*B*	*B*−	*C*+	*C*	*C*−	*D*+	*D*	*D*−	*E*
MALE	573	144	189	153	157	179	164	184	166	326	351	305
FEMALE	551	223	235	180	198	201	194	236	184	406	338	298

χ^2 = 29.358 df = 11 P-Value < 1.997×10^{-3}

In the table above (Table 4.20) on the performance in mathematics, out of the sample of 6135 students. The grades were distributed as follows: 6% of the sampled students scored B+, B-, C+, C and D+ respectively, 18% scored A, 7% B+ & C- respectively, 5% B, 12% D, 11% D- and 10% E.

Table 4.21. English performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	154	271	430	433	409	347	238	177	133	215	78	6
FEMALE	64	191	581	757	635	382	213	121	87	172	39	2

χ^2 = 234.15 df = 11 P-Value < 2.2×10^{-16}

Table 4.21 with the performance in English shows that: 4% of the sampled students scored A and D+, 8% A-, 16% B+, 19% B, 17% B-, 12% C+, 7% C, 5% C-, 6% D, 2% D- and those who scored a grade E in English were less than 1%.

Table 4.22. Swahili performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	373	250	292	257	268	342	270	265	155	298	103	18
FEMALE	332	351	420	300	342	415	344	251	164	251	67	7

χ^2 = 67.658 df = 11 P-Value < 3.397×10^{-10}

Table 4.22 shows that 11% of the sampled students scored A,10% A- , C & B- respectively, 12% B+ & C+ respectively, 9% B & D respectively, 8% C-, 5% D+, 3% D- and those who scored E in swahili were less than 1%.

Table 4.23. Biology performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	381	176	195	263	274	284	206	235	164	402	236	75
FEMALE	55	115	240	361	378	441	328	345	243	440	235	68

χ^2 = 374.93 df = 11 P-Value < 2.2×10^{-16}

Table 4.23 on the performance in Biology by the 6135 sampled students had grades distributed as follows:7% of them scored A, B+ ,C+, & D+ respectively, 5% A-, 10% B, 11% B-, 9% C & C- respectively, 14% D, 8% D- and 2% scored grade E.

Table 4.24. Chemistry performance.

GRADE / GENDER	A	$A-$	$B+$	B	$B-$	$C+$	C	$C-$	$D+$	D	$D-$	E
MALE	322	137	151	217	229	335	244	251	181	427	295	102
FEMALE	227	151	207	332	239	337	261	315	263	510	325	77

χ^2 = 65.342 df = 11 P-Value < 9.292×10^{-10}

The table above (table 4.24) with the distribution of the grades scored in chemistry by the 6135 sampled students: 9% of the students scored A, B & C- respectively, 5% C-, 6% B+, 8% B- & C respectively, 11% C+, 15% D, 10% D- and 3% E.

Table 4.25. Summary of school category and number of students with reference to gender.

SCHOOL CATEGORY	FEMALE	MALE	TOTAL
National	615	808	1423
Extra County	604	1154	1758
County	694	627	1321
Sub County	892	502	1394
Private	86	153	239
TOTAL	2891	3244	6135

χ^2 = 310.25 df = 4 P-Value < 2.2×10^{-16}

Table 4.25 shows that out of the 6135 students sampled, 23% were from national schools and out of that 43% were females and 57% females. 29% were from Extra county schools out of which 34% were females and 66% males. 21% were from County schools out of which 53% Females and 47% males. 23% were from sub county schools where 64% were females and 36% males. 4% of the sampled students were from private schools out of which 36% were females and 64% males.

Table 4.26. Summary of school category and overall grades attained.

SCHOOL CATEGORY	A	B	C	D	E	TOTAL
National	804	534	65	19	1	1423
Extra County	180	977	564	37	0	1758
County	7	282	781	249	2	1321
Sub County	3	175	501	703	12	1394
Private	13	96	69	56	5	239
TOTAL	1007	2064	1980	1064	20	6135

χ^2 = 4399.3 df = 16 P-Value < 2.2×10^{-16}

Table 4.26 on the summary of school category and overall grades scored shows that 98% of the grade A scored came from national and extra county schools while 2% came from county, sub county and private schools. 73% of the grade B scored were from national an extra county schools wile 27% were from county, sub county and private schools. 32% of the grades C were from national and Extra county schools while 68% were from county, sub county and private schools. 5% of the grade D scored were from national and extra county schools while 95% were from county, sub county and private schools. Only 1 student from a national school scored a mean grade of E, 2 from county, 12 from sub county and 5 from private schools.

Table 4.27. A comparison between the subjects that contribute the most to students overall performance interms of odds ratio for both the years 2014 and 2015.

2014		2015	
SUBJECT	ODDS RATIO	SUBJECT	ODDS RATIO
Biology	3.7933	Swahili	3.3624
Swahili	3.656	Chemistry	2.9886
Chemistry	3.2552	English	2.9640
English	3.1923	Mathematics	2.8539
Mathematics	2.7174	Biology	2.7461

Table 4.27 above showing the subjects in order of merit of how much they contribute to the overall performance of students interms of odds ratio for both the year 2014 & 2015.

Swahili and Chemistry contribute the most out of the 5 subjects used as predictors: Swahili had an odds ratio of 3.3624 in 2015 and 3.656 in 2015, Chemistry had an odds ratio of 3.2552 in 2014 and 2.9886 in 2015. Mathematics contribution is among the least, in 2014 it

had an odds ratio of 2.714 (the least contributing subject) and in 2015 it had an odds ratio of 2.8539 (the second last in order of odds ratio). Biologys' contribution is not consistent, in 2014 it contributed the most with an odds ratio of 3.7933 and in 2015 it contributed the least with an odds ratio of 2.7461. The contribution of English subject is average, in 2014 it was 4^{th} with an odds ratio of 3.1923 and in 2015 it was 3^{rd} with an odds ratio of 2.9640.

5 CONCLUSION AND RECOMMENDATIONS

5.1 CONCLUSION

On the data analyzed for year 2014.
Biology seemed to lead with an odds ratio of 3.7933. It contributed the most to the overall performance of the learners. It was followed by Swahili with an odds ratio of 3.6256, which was followed by Chemistry, English, and mathematics, that had odds ratios of 3.2552, 3.1923 and 2.7174 respectively.

Students who attended National Schools were more likely to perform better, followed by those who attended Extra County Schools. County schools came third, Sub-county schools were forth then private schools.

In the year 2014, gender was not significant.

Reject the 1st Null Hypothesis since Biology is a science and the highest contributor (As indicated by the high odds ratio of 3.7933).

Fail to reject the 2nd Null Hypothesis as there is no correlation between students gender & their performance as was indicated by the P-value of 0.3081 which is greater than α (0.05).

Reject the 3rd Null Hypothesis as there is a difference in the performance of students in the different categories in schools.

On the data analyzed for the year 2015.
Swahili contributed the most to the students' overall performance with an odds ratio of 3.3624. It was followed in that order by Chemistry, English, Mathematics and Biology with odds ratios of 2.9886, 2.9640, 2.8539 and 2.7461 respectively.

Gender was significant where female students were more likely to perform better than their male counterparts.

Students from National Schools were most likely to perform better than those from Extra County schools, then county schools, sub-county schools and finally students from private schools.

Fail to reject the 1st Null Hypothesis since Swahili contributes the most to the students overall performance (As indicated by the high odds ratio of 3.3624).

Reject the 2nd Null Hypothesis as there is a correlation between students gender & their performance as was indicated Female students are are 2 times more likely to perform better than male students.

Reject the 3rd Null Hypothesis as there is a difference in the performance of students in the different categories in schools.

5.2 RECOMMENDATIONS

The MOEST should come up with mechanisms of making language subjects such as Swahili and English more popular among students because they have a significant contribution to students' overall performance in national exams.

Since the girl child has been empowered more than the boy child, it is time to give the same empowerment to the boy child because this is likely to significantly improve the male students' performances.

Upgrading, equipping and allocation of more funds to county, sub-county and private schools as this may contribute to improved performance among students in these schools.

5.3 AREAS FOR FURTHER RESEARCH

- ✓ More research should be done on the reasons as to why the performance of male students is deteriorating.
- ✓ Research should be done on the reasons/ causes of poor performance among students from private schools.
- ✓ Research should be done on the impact of teacher-student ratio on the overall performance of students.
- ✓ Research should be done on the contribution of optional subjects such as Physics, Humanities, Technical subjects and Foreign languages among others, on the overall performance of students.

Bibliography

[1] Amunga .J.k,Amadalo M.M & Musera .G. (2009). Disparities in chemistry and biology achievement in secondary schools : Implications of vision 2030. *International journal of humanities and social sciences*.vol 1 number 18,pp 226-236.

[2] Anyango,H.O (2001). Influence of headteachers' leadership styles on students KCSE performance in Mombasa District, *Unpublished MEd Thesis, Kenyatta University*,Nairobi.

[3] Asena M.J,Aggrey M.S,Andrew R. (2016) Factors affecting subsidized free day secondary education in enhancing learners retention in secondary school.*journal of education and practice*.vol 7 number 20.

[4] Cox,D.R and Snell,E.J (1989) Analysis of Binary data.Second edition. Chapman & Hall.

[5] Cynthia.N.K(2014). Factors influencing Academic performance of students in KCSE: A case study of Roysambu constituency, Nairobi Kenya.

[6] Daniel K.K & Beatrice C.J (2014). An assessment on the differences in performance In mathematics & chemistry between boys & girls in secondary schools in Bomet District,Kenya. *International journal of advanced research (2014)*,volume 2,issue 2,1348-1355.

[7] Das.s and Rahma,M.R (2011). Application Of Ordinal Logistic Regression Analysis in determining risk factors in child malnutrition.

[8] Elvis.K.M,Nelson.O.O (2015). Analysis of KCSE performance in Nakuru County: A generalized Estimating Equation Approach.*Applied mathematics,* 2015,6,2217-2225

[9] Eshiwani,G.S (1993). Education and Development in Africa . London: The MaCmillan Press Limited.

[10] Eshiwani,G.S (1993). The education for women in Kenya. *Bureau of Education Research*, Kenyatta University .Nairobi.

[11] Githua,B.N,(2001); Factors Affecting Teaching and Learning of Mathematics in Kenya Secondary Schools;Unpublished Paper Presented to Nakuru District Science and Mathematics Teachers Seminar,Nakuru High School, 9th-12th April,2001.

[12] Glassman, D., & Sullivan, P. (2006).Governance, Management, and Accountability in Secondary Education in Sub-Saharan Africa. *Human Development Department, Africa Region (AFTHD)*, World Bank.

[13] Hosmer,D.W,Jr,S.A Lemeshow and Sturdivant,R.X (2013) Applied Logistic Regression.3rd e.d Hoboken,Nj:Wiley.

[14] Hudson,H.T and Rottoman,R.M (2006). Correlation Between the Performance in Physics and Prior Mathematics Knowledge. *Journal of Research in Science and Technology.* Vol. 18 issue number 4.

[15] Hutchings, T.L (1973). Teaching Mathematics at Secondary School.Cambridge University Press,London.

[16] Jane K.A,Musasia M.A,Geofrey M.(2011); Disparities in physics Academic achievement and Enrollment in secondary schools in western province;implications for strategy renewals;*problems of education in the 21st century,*vol 31,2011 pp 18-22.

[17] Keith,M. (1996); Introduction to High School Physics 2nd edition.*Oxford University Press*.

[18] Kleinbaum,D.G and Klein,M.,(2010); Logistic Regression. A self learning text.Third edition,Springer.

[19] Kenya National Examinations Council (1995); The 1993 and 1994 KCSE Mathematics Report.

[20] Kenya National Examinations Council (2000); The KCSE.News letter Nairobi: KNEC.

[21] Kenya National Examinations Council (2002); The 1989 – 2001 KCSE Mathematics Performance Analysis Report;Nairobi:Unpublished.

[22] Kenya National Examinations Council (2004 – 2010); KCSE Examination Report: KNEC, Nairobi.

[23] Lewis M.N & Esther N.K(2014); A study of students' performance in KCSE with special reference to biology.*Journal Of Education Research*, paper 17.pp 191-201.

[24] Lydia M.M, Joash .M & James M.M(2013); Determinants of girls' performance in science, mathematics & technology subjects in public secondary schools in Kenya *International journal of education administration and policy studies*, vol.5(3) pp 33-42.
http://www.academicjournal.org/IJEAPS.

[25] Lyons,T.(2006); The Puzzle of Falling Enrollment in Physics and Chemistry Courses: Putting Some Pieces Together. *Research in science education*, 36(3), 285-311.

[26] Mc Fadden,D. (1973); Conditional logit analysis of qualitative choice behavior PP 105-142 in P. Zaremka (e.d). *Frontiers in economics*., Academic press.

[27] Mwangi,D.T (1983). A Study of the Variation of Teaching Mathematics at Different Class Levels in Selected schools of Mombasa District. M.Ed Thesis,Kenyatta University.

[28] Nderitu,M.K (2011); Determinants of Enrollment and Performance in Physics in Sellected Secondary Schools in Murang'a District.Unpublished M.ed Thesis.

[29] Ngware,M.W., Ciera,J.,Musyoka,P.K & Oketch,M(2015); Quality of Teaching Mathematics and Learning achievement gains : Evidence of Primary Schools in Kenya.*Educational Studies in Mathematics*, 89(1), 111-131
DOI 10.1007/510649-015-9594-2.

[30] O'Connel,A.A (2006). Logistic regression models for ordinal response variable; Quanitative Application in social sciences,Sage publication. Inc

[31] Ombui,G.M & Gichuhi,A.W (2011). Using Ordinal Logistic Regression Modeling to Evaluate the Sartisfaction of Jomo Kenyatta University of Agriculture & Technology faculty of science students. *Journal of Agriculture, Science and Technology*, vol 13(1).

[32] Onyango,J.A (2012). Factors That Influence girls' Performance in Mathematics in Kisumu City Public Secondary Schools' ,M.Ed Thesis, Kenyatta University,Nairobi ,PP ii.

[33] Orodho,J.A (1996). Factors Determining achievement in Science Subject at Secondary School Level In Kenya.Unpublished PhD dissertation, Kenyatta University.

[34] Orodho,J & Njeru.E. (2003). Education Financing in Kenya: The secondary school busary scheme.*Institute of Policy Analysis and Research-Discussion Paper*,(35). Rieta Printer. Nairobi.

[35] Orodho,J. & Kombo, W.(2004). Educational Research Methods.*Rieta Printer. Nairobi*.

[36] Shakhawat.H,Ahmed.S.E and Hatem.A.W (2014) Model selection and parameter estimation of multinomial regression model.*Journal of statistical*

computation and simulation.84:7, 1412-1426,
Doi 10.1080/00949655.2012.746347.

[37] Shikuku,B.N. (2009) Effect of Syllabus Coverage on Student Performance at KCSE Mathematics: A Case of Kakamega South District,Kenya.*lap Lambert Academy Publishing*: rehagmbh,DudwilesrstraBe 7266111 Saarbrucken. www.rehagmbh.de.

[38] UNESCO, (2005). Innovation: applying knowledge in development. Paris: United *nations Millennium Project task force.*

Appendices

Appendix A: MAP OF KIAMBU COUNTY

This is the map of Kiambu County showing all the 12 sub counties.

Appendix B: R CODES USED

```
>rm(list = ls())
>require(foreign)
>require(ggplot2)
>require(MASS)
>require(Hmisc)
>require(ordinal)
>require(reshape2)
>=read.csv(file.choose(), header = T, sep = ",")
>attach(kcse2014)
>View(kcse2014)
```

Analysis for the year 2014

```
>kiambucountykcse2014<-polr(as.factor(MEANGRD)~MATHS+ENG+KIS+BIO+CHEM
+as.factor(GENDER)+as.factor(SCHLCATE), Hess = T);kiambucountykcse2014
>summary(kiambucountykcse2014)


## checking the structure of the data
>str(kiambucountykcse2014)

## Two way table of factor variable
>xtabs( GENDER + SCHLCATE, data = kcse2014)
>xtabs( GENDER + SCHLCATE, data = kcse2015)


## store table
>(ctable=coef(summary(kiambucountykcse2014)))


## calculate and store p values
>p=pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2


## combined table
>(ctable=cbind(ctable, "p value" = p))
```

```
## OR and CI
>exp(cbind(OR = coef(kiambucountykcse2014), ci))

## running a crosstabulation or chi square test.
>table(SCHLCATE,CHEM)
>CHI<-chisq.test(table(SCHLCATE,CHEM));CHI
>summary(CHI)
```

Analysis for the year 2015

```
## ANALYSIS OF 2015
>rm(list = ls())
>kcse2015=read.csv(file.choose(), header = T, sep = ",")
>attach(kcse2015)
>View(kcse2015)

>kiambucountykcse2015<-polr(as.factor(MEANGRD)~MATHS+ENG+KIS+BIO+CHEM
+as.factor(GENDER)+as.factor(SCHLCATE), Hess = T)
>summary(kiambucountykcse2015)

## store table
>(ctable=coef(summary(kiambucountykcse2015)))

## calculate and store p values
>p=pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2

## combined table
>(ctable=cbind(ctable, "p value" = p))

>(ci <- confint(kiambucountykcse2015)) # default method gives profiled CIs

## odds ratios
>exp(coef(kiambucountykcse2015))

## OR and CI
>exp(cbind(OR = coef(kiambucountykcse2015), ci))

## running a crosstabulation or chi square test.
>table(SCHLCATE,MEANG)
>CHI<-chisq.test(table(SCHLCATE,MEANG));CHI
>summary(CHI)

>confint.default(kiambucountykcse2015) # CIs assuming normality
```

odds ratios (exponentiating the coefficients so as to interpret them)
>exp(coef(kiambucountykcse2015))
>exp(coef(kiambucountykcse2014))

OR and CI
>exp(cbind(OR = coef(kiambucountykcse2014), ci))
>exp(cbind(OR = coef(kiambucountykcse2015), ci))

Testing Various Assumptions

testing the proportional odds assumption ##2014
>sf <- function(y) {
c('Y>=1' = qlogis(mean(y >= 1)),
'Y>=2' = qlogis(mean(y >= 2)),
'Y>=3' = qlogis(mean(y >= 3)),
'Y>=4' = qlogis(mean(y >= 4)),
'Y>=5' = qlogis(mean(y >= 5))) };sf
>(s <- with(kcse2014, summary(as.numeric(as.factor(MEANGRD)) ~ MATHS+ENG+KIS+BIO+CHEM
+as.factor(GENDER)+as.factor(SCHLCATE), fun=sf)))

testing the proportional odds assumption ## 2015
>sf <- function(y) {
c('Y>=1' = qlogis(mean(y >= 1)),
'Y>=2' = qlogis(mean(y >= 2)),
'Y>=3' = qlogis(mean(y >= 3)),
'Y>=4' = qlogis(mean(y >= 4)),
'Y>=5' = qlogis(mean(y >= 5)))};sf
>(s <-with(kcse2015,summary(as.numeric(as.factor(MEANGRD))~MATHS+ENG+KIS+BIO+CHEM
+as.factor(GENDER)+as.factor(SCHLCATE), fun=sf)))

>glm(I(as.numeric(as.factor(MEANGRD)) >= 5) ~ MATHS+ENG+KIS+BIO+CHEM+as.factor(GENDER)
+as.factor(SCHLCATE), family="binomial", data = kcse2015)
>summary(kiambucountykcse2014$sf)

TESTING NORMALITY
>library("ggpubr")
>ggscatter(kcse2014, x = "ENG", y = "CHEM", add = "reg.line", conf.int = TRUE, cor.coef
= TRUE, cor.method = "pearson", xlab = "Miles/(US) gallon", ylab = "Weight (1000 lbs)")
>shapiro.test(kcse2014$ENG)

```
>library("ggpubr")
>ggqqplot(kcse2014$SCHLCAT, ylab = "SCHLCAT")

## Goodness of fit test
with(kiambucountykcse2014,pchisq(null.deviance-deviance,df.null-df.residual,lower.tail =
F))
```