

Mining Educational Data to Analyze Students' Performance: A Case Study of Mawuli School, Ho

Cephas Lokpo
Mathematics/ICT Department
Akatsi College of Education
Akatsi, Ghana

Abstract:- Data mining is the extraction of prospective valuable information from large chunk of data through the employment of many different data mining techniques. The usefulness of data mining coupled with the huge data generated in scholastic settings has made it an interesting field of research known as Educational Data Mining (EDM). The intent of EDM is to derive understanding from hidden patterns in data collected from institutions of learning to aid in identifying issues that influence students' scholastic accomplishment, to solution of which will lead to improvement in accomplishment. Because scholastic achievement is dependent on several issues, it is essential to develop predictive models on students' academic performance. This study's objective, therefore, is to acquire an insight into performance through knowledge discovery by the use of simple linear regression in order to build a predictive model capable of predicting students' grades to give a general overview of students' performance in the WASSCE, and help improve students' performance. In accomplishing the set objective for which the study was carried out, that is to predict the possible outcome of students in WASSCE, a widely sampled study was applied in undertaking the study (Quantitative research). The case study chosen was Mawuli School Correlation and Regression was employed to determine the degree to which variables are connected and to construct a model to predict students' final scores. It was revealed that there was a decline in students' performance in Mathematics, English Language, and Integrated Science, and a good number of students fell below the pass mark, although some few students were performing high, while Social Studies saw a good improvement in students' performance over performance in the first year by recording a 98.9% pass. It was also discovered that there is a close association between students' first year and second year performance, and that the effect of students' first year performance on their second year performance was not due to likely random chance but that the effect had underlying repeatable cause, and hence models were constructed to predict students' final WASSCE scores/grades.

Keywords:- Educational data mining, performance, association, prediction.

I. INTRODUCTION

Predicting students' scholastic accomplishment is critical for monitoring progress in establishments of learning. The scholastic output of students is centered on some varying issues namely psychosomatic, social, personal and other environmental agents. Data Mining is a very useful tool that can be employed in helping to achieve objectives of scholastic accomplishment. Tools used in extracting knowledge have the capability of maneuvering volumes of sets of data to determine unseen samples besides being helpful links in decision making (Bharadwaj and Pal, 2011). Prediction models have a definite goal that enables us conjecture unspecified estimates of the observed variables given recognized values of other variables. In predicting the accomplishment of the learner effectively, it is necessary to use plans such as the earlier mentioned variable elements. The benefit of predicting the accomplishment of students with precision helps in identifying students who perform low academically in the beginning, and the subject areas that students need further assistance. It is essential for the teacher to give more assistance to such students to help ameliorate their scholastic accomplishment in the future.

The integration of IT in diverse disciplines of human endeavor has led the abundant storage of data in data repositories in different formats. Data collection from diverse applications necessitates appropriate technique for digging up information from a vast data warehouse for informed decisiveness. Mining of data, also known as Knowledge databases discovery or (KDD), focuses on discovering valuable information in large sets of data. Furthermore, the Core tasks of mining of data include utilizing diverse techniques as well as processes which is to uncover and dig up models of amassed data.

The process of extracting informative knowledge which includes samples, relations, variations, irregularities and considerable organizations obtained from vast distribution of information deposited in data warehouses, databases, or any data storehouses is referred to as Data mining (DM). (Han and Kamber, 2001). It is regarded as one of the knowledge discovery steps which interrelate with knowledge or user base. Patterns of interest are offered to users, or maybe kept as fresh information in the database. Even though it is just one stage of the discovery of knowledge processes it is very important since unseen models are discovered for assessment (Han and Kamber,

2000). Educational Data Mining (EDM), deriving knowledge from patterns and trends uncovered from learning institutions is a field of research of contemporary time which is concerned for fashioning devices that uncover information from educational databases (Han and Kamber, 2000).

These records can be retrieved from archival and active data catalogues in educational institutes. These collected data are either of educational or individual's particulars. Furthermore, e-learning provides large volumes of data that can be utilized (Romero and Ventura, 2007).

Analyzing data gathered from scholastic establishments through the use of DM systems to solve challenges associated with research in education is termed EDM. The purpose of undertaking EDM is to have a better knowledge of students and their place of learning. Therefore, EDM's center of attention is gathering, storing and interpretation of facts and figures from students' studies and evaluation.

Various methods are applied in undertaking EDM which include Naïve Bayes, Decision Trees, Nearest neighbor, Neural Networks, K- Regression, Correlation, etc.

Employing methods like classifications, clustering and association rules can discover a lot of information that would be useful in predicting student enrolment for a given course, detecting electronic exam malpractices, errors in student reports, performance of students and many more. From the perspective of learning, standardized methods for measuring the ability and qualities of the human mind which lack precision can find a replacement of EDM. At present, a good deal of EDM study areas well accommodated in the environment of Intelligent Tutoring System (ITS) because of the greater number of researchers in the EDM domain focusing attention on mining data originating from ITS.

EDM searches for a better understanding of the learning process and the participation of students of it, always searching for a quality improvement and cost-effectiveness of the education system. Educational Data Mining has the following objectives (Romero and Ventura, 2007):

- Pedagogical: to help in the design of didactic contents and learners' scholastic accomplishment promotion;
- Managerial: optimization of the establishment and preservation of structures of training;
- Commercial: enrollment of students in institutes of learning that are owned by individuals nongovernmental groups in particular.

A. Statement of Problem

In Ghana's educational system, students' performance in the senior high school is defined by summing up of the marks obtained from examination set by an independent body and marks accumulated from terminal exams and class work. A student is successful when he or she excels above the minimum academic requirements both in school and in WASSCE.

The grades of students based on performance are not in any way predictable in this modern system of education. Most of the renowned second cycle institutions in Ghana, and most especially in the Volta Region of Ghana, have seen a serious decline in performance in recent times. The results of 2013 revealed a sharp fall in performance. For instance, in one of such schools, out of 1089 candidates who took the examination, 599 students passed in all eight subjects, representing 55% and only 471 qualified for tertiary level, representing 43.5%. This decline is so sharp when compared to 2011 and 2012 results that recorded 90.6% and 89.3% of candidates that qualified for tertiary level.

In spite of this situation, stakeholders always expect high performance even though the stock of candidates presented for WASSCE may not be the best, and most often results of poor performance hit them by surprise. Perception among some stakeholders is that the stock of students being admitted through the Computerized School Selection and Placement System (CSSPS) to a large extent is affecting performance negatively.

B. Purpose of the Study

The study intended to gain insight into students' performance through knowledge discovery in order to establish a predictable system capable of predicting students' grades in WASSCE in order to give an overview of the general performance of students and better inform management and other stake holders with regard to their decisions and expectations. This will also help identify how weak or high students' grades may be. These projected grades will help determine the amount of remedial efforts that need to be expended by teachers to help students improve their performance before their final examination to reduce failure rate.

C. Objectives of the Study

1. Analyze the performance of students in the subjects namely Integrated Science, Mathematics, English Language, and Social Studies, using linear regression.
2. Verify the extent to which the grade of students in Basic Education Certificate Examination (BECE) is connected to the grade of students in Senior High School (SHS).
3. Verify the level of connection between first and second year performances of students in all core subjects.
4. Construct a predictive model for projecting students' final grade in all core subjects (Integrated Science, Mathematics, English Language, and Social Studies) using linear regression analysis.

D. Research Questions

1. Do all students perform above average (50%) in Integrated Science, Mathematics, English Language, and Social Studies?
2. Do students' grades in junior high school exam (BECE) determine their performance in senior high school and to what extent?
3. Do students' terminal performance at SHS determine their final WASSCE grades and to what extent?

E. Significance of the Study

Undertaking DM proffers extensive accurate tools for making decisions, When applied in education, DM tasks indubitably offer broad, yet precise, decision-making tools, detection as well as projection of the consequence of gaining knowledge and response, and learners' rating forecasting. Others include detection of learners who have the possibility of not being retained, learner counseling; recognition of learners who are excelling, poor-performing students, learners with the same learning characteristics and way of thinking, disagreeable detecting learner demeanor detection, recognition of learners who are apt to team work, assigning the right learning task to the right group of learners, preparation and drawing timetable for lessons and other learning behaviors, etc. (Sargenti et al., 2006; Rajagopalan and Shyamala, 2006).

In view of the above quote, this research will in no small way help in resolving the challenge of low performance of students because the predictions will reveal the nature of the final result and the subject areas where more attention must be given. The highest, lowest, and the average marks predicted by the regression model speak a lot about the overall performance of students. The study is therefore relevant to management and teaching staff of Mawuli School for decision making, as it gives the overview of students' performance in their final examination. Stakeholders will get the general overview of students' projected final results.

II. LITERATURE REVIEW

Interests have grown in mining educational data and it is all about formulating techniques to unearth information from educational settings (Garcia et al., 2008).

Data is made available in archives (historic data), and from information systems on day-to-day transactions (operative data) found in repositories of educational establishments. Information of students gathered on demographic variables (personal data) or on performance variables (academic data). Similarly, electronic systems of learning provide such information (Romero and Ventura, 2007).

Higher educational establishments seek to offer its students with education of the highest quality and advance their decision-making ability. And this can be achieved by unearthing information from their academic data to evaluate the issues that influence learners' scholastic accomplishment. The extracted knowledge is beneficial in contributing to decision making in the administrative level as it is productive reference decision makers. Some benefits it provides include improving the academic output of students, reduce poor performance, good comprehension of student conduct and improve the overall educational process (Chadha and Kumar, 2011). Educational data mining utilizes numerous methods in discovering data employing techniques for instance association rules, clustering and classification. These include decision tree, rule induction, neural networks and so on (Romero and Ventura, 2007).

A. Mining of Data

The discipline associated with the discovery of possibly valuable data from a vast distribution of data is called data mining or Knowledge Discovery in Database (KDD). Fields of application include marketing, fighting terrorism and management of biological information. The urge to undertake scientific investigations in the field of EDM has soared up considerably in contemporary time. Of late attention is drawn to the utilization of techniques used to mine data to examine technical problems in research areas in education; this is commonly called Educational Data Mining.

The term used to describe a rise in the volume of information stored in an establishment's database as well as probability to uncover depths in analyzed is "big data." According to IBM, "big data" covers three sundry scopes, namely volume, velocity, and variety (IBM, 2012).

Filtering to discover useful information for decision making is a very big challenge to organizations. One efficient instrument to help organizations a great deal to unearth important information for decision making at all levels of management operations is data mining (Kiron et al., 2012).

Mining of Data involves the use of categories of tools and techniques to uncover unrecognized samples and associations found in the data (Dunham, 2003). DM similarly is by and large a single stage in the process of discovering information, in which institutions desire to be well informed in taking decision using fresh data. It can therefore be said that in taking major decisions, DM as well as discovery of information plays a crucial and inevitable role.

Employing diverse methods of DM like classifications, clustering and association rules can be used to discover a lot of information, as such techniques measurably evaluates enormous amount of information to discover unseen designs and interpretations of information. DM therefore confirms evaluations as well as explains the process (Berson et al., 2011). Being highly exploratory, DM varies from investigative and analyzing procedures, as it is typically determined by the problem and confirms.

Areas where DM had is employed in various organizations, in defense as well as in commerce however, it has been underutilized in the area of education. (Malik and Ranjan, 2007). The story is different today; however, as data mining has seen a significant application in the field of education. This buttresses the point Chakrabarti and others made that in recent times that data mining is now greatly considered both in education and business (Chakrabarti et al., 2009).

DM in education is an aspect of evaluating and applying DM in solving challenges pertaining to learners and their settings in which they learn (educationally-related problems). Modern techniques in unearthing designs as well as developments in diverse sets of information in education

is made available to experts in the research field when DM is applied in this way.

From the ongoing discussion on concepts of data mining, it can be deduced that it is a process of information extraction from a mass of knowledge, in agreement with (Han and Kamber, 2001), which stated that: DM is a method of unearthing fascinating information deposited in huge storage systems of any kind such as a repository of data, data store houses. Fascinating data as in one unseen by any normal technique but of immense importance.

Considering the import of the definition and explanation of data mining by the various authors, all of them seemed to be dwelling on the same idea. The pivotal issue in all definitions is the discovery of valuable information from patterns, trends, and relationships extracted from huge or complex data sets with a primary purpose of making informed decisions that will improve efficiency.

Data mining encompasses various fields of computer science. The problem at hand which require a solution as well as the trend of data of interest, are the factors that determine the type of data mining technique to use. However, as determined by the situation at hand numerous procedures could be approached such as learning aided by machines (Witten and Frank, 1999); man-made intellect, data extraction and organization (Mitra and Acharya, 2003); and many more. It can also be said that, various discovering strategies yield different trends and models of data these may include analyzing in clusters, analyzing by trend, etc. (Han and Kamber, 2001).

By a careful examination of what techniques, one can employ in data mining it is obvious from the previous paragraph that there are many techniques as suggested by (Mitra and Acharya, 2003; Chakrabarti et al., 2009). The choice of particular technique however, is determined by the kind of problem the researcher is seeking to solve and the type of pattern and data to be mined as proposed by (Witten and Frank, 1999). In a similar manner, (Han and Kamber) suggested that various discovering strategies yield different trends and models.

The general means of discovering knowledge by the use of data mining techniques involves an interactive order as follows: i) Preprocessing step which involves data cleaning and amalgamation ii) modeling step which involves selecting data, transforming it, structuring, trend valuation, and information presentation (Han and Kamber, 2006). Data is analyzed for the purpose of removing noise, missing values and inconsistencies before the processing stage. The resultant preprocessed is deposited in data storage bases. Several approaches for example, recognizing of patterns 'dig out' and assess configurations in data during the modeling step. The relevant trends discovered could be delivered to customers through visualizing and presentation methods (Chakrabarti et al., 2009).

Stages in the KDD method are represented below:

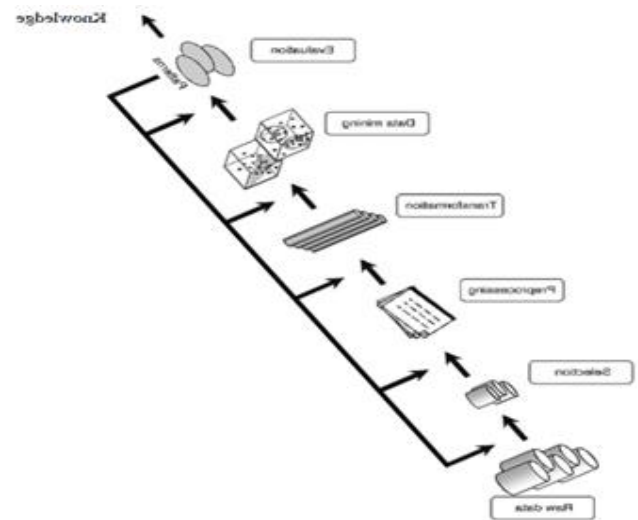


Figure 2.1: DM: KDD procedure (The steps of extracting knowledge from data)

Source: Baradwaj and Pal (2011)

The following are areas suitable for data mining:

- Prediction- estimate verification of unknown attribute on the basis of designs inherent in the data.
- Classification- categorization of facts and figures according to characteristics.
- Clustering- identification of what is similar and dissimilar in features of a set of data for discovering homogeneous/heterogeneous groups that will portray the records.
- Description- interpretation of patterns in the data into forms that human beings can understand (Fayyad, 1996).

DM functions similarly to the brain of humans. It learns new knowledge based on historical information or experience. That notwithstanding, for data mining technology to arrive at new knowledge (information) in the database, the user must present to it the patterns. DM utilizes these pieces of data from the past to generate a model that will predict what may occur with time by using such information characteristics to identify similar ones in data storage.

To build predictive or inference models whose aim is to forecast or project future tendencies or behaviors established on the examination of prearranged data is a function of data mining (Han and Kamber, 2001). Prediction as used contextually means the building of structures that is used to evaluate the kind of unidentified instance. Alternatively, it means to examine or analyze the estimate or value ranges of a character trait that a given instance probably has. Two mainly used predicting methods are Classification which is exercised in predicting distinct figures and regression which is applied in predicting well-ordered data constitute two principal techniques for forecasting. Classification is the method employed in predicting discrete or nominal estimates and method of regression utilized for projecting constant estimates (Larose, 2006).

B. Educational Data Mining

Educational data mining (also referred to as “EDM”) is defined as the area of scientific inquiry centered on the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to understand students and the settings which they learn better.

Educational data mining methods often differ from methods of the broader data mining techniques in literature. They are methods explicitly used for exploiting the multiple levels of meaningful hierarchy in educational data. To achieve this goal, there is often the integration of methods from the psychometrics literature and methods from the machine learning and data mining literatures. An example to this effect is the concurrent consideration of data generated at the keystroke level, answer level, session level, student level, classroom level, and school level when examining the choice of students in the use of educational software. Issues of time, sequence, and context also play important roles in the study of educational data. With the establishment of the annual International Conference on Educational Data Mining, and the Journal of Educational Data Mining in 2008, an independent area of research has ever since come to light.

Different definitions exist on EDM by various authors. (Campbell and Oblinger, 2007) defined the analysis of data gathered from scholastic establishments to mean the amalgamation of methods used for mining/analyzing data in the field of statistics to facilitate the promptness of staff as well as counselors/tutors to identify learners who are likely to fail their assessment so that appropriate remedial action can be taken. By this, the probability of learner drop out is reduced through the discoveries of DM. Analysis of scholastic data centers attention on course of actions which took place at various levels like institution, department, school, and class. Data analysis of this sort leaves out further information about every subject, and for that matter it gives a general description. It comes under Educational Data Mining.

Evaluation of the two definitions points to a fact that academic analysis of academic data has to do specifically with data associated with good organization of learning establishments as well as drop out of learners.

(Baker and Yacef, 2009) defines Educational Data Mining as a developing field of research that generates techniques capable of analyzing distinct educational information from academic institutions and utilizing the information to appreciate better student behavior and their academic environment. This definition left out data mining being mentioned, giving room to researchers to explore and develop further means of analysis that can be relevant to educational data. One advantage of their research is the broad representation of the EDM field it provides by thrashing out issues in the outstanding research reports. Their research however used a number of article citations as a means of evaluating/appraising the growth of EDM.

Perhaps in future a broader perspective can be used in researches that seek to evaluate the growth of this discipline.

EDM spans spheres of dramatic influence on learners. Further themes that EDM encompasses have to do with analyzing didactic procedures such as enrollment, old student associations, and choice of program. In addition, there are unique tools for mining data gathered from academic institutes and can be functional in extraction of data online, categorization, mining to discover rules of connections, and analysis of data that consist of many variables (Calders and Pechenizkiy, 2012). These methods of analyzing data are to very large extent investigative techniques and are used for estimating and projecting learning outcomes and improvement needs of institutions. In addition, dissimilar personal behaviors can be made a prototype so that the right steps to deal with such dissimilar behaviors for the improvement of acquisition of knowledge (Corbett, 2001).

A solid data warehousing strategy is essential for a successful educational data mining. The indispensability of applying relevant information in taking decisions in learning establishments has been emphasized by (Guan et al., 2002). The challenge then is how to acquire quick and efficient data to meet the needs of decision makers. There are main factors that kick off data warehouse projects. These include improved environment for competition, large task of passing information to stakeholders, namely community leaders, board members, parents and legislators.

Ideas from organizational data mining can be used in EDM. In mining data generated from the business environments attention is centered on how establishments can be assisted to maintain competitive advantage (Nemati and Barko, 2004). The reliability of ODM on organizational theory as a field of reference is what makes it different from EDM (Nemati and Barko, 2004). The benefits that organizations can reap if they are able to transform their data efficiently into valuable information and knowledge are tremendous. Such benefits include augmented competitiveness, improved decision-making, and prospective financial growth/increase (Nemati and Barko, 2004).

EDM centers attention primarily on quantitative analyses. To back case researches in EDM, interviews and document analysis which are qualitative techniques are used. Quantitative research remains the principal research paradigm of EDM with associations, predictions, classifications, and clusters as forms of corresponding results. The disadvantage of done in EDM is that outcome of such studies cannot be generalized for establishments within similar domain. What it implies is, outcomes extremely relate to an exclusive institution at a definite time.

C. Approaches of Mining Educational Data

Within EDM, there exists an extensive diversity of modern renowned methods. The general categories are forecasting, homogeneous / heterogeneous grouping, association discovery, using prototypes to extract knowledge, and filtering raw facts. Prediction, clustering, and relationship mining are widely recognized across types of DM while prototypes to extract knowledge and filtering raw facts are forms that particularly obtain eminence within EDM.

(Baker and Yacef, 2009; Baker, 2010) suggests four key areas of application for EDM: improving student models, improving domain models, studying the pedagogical reinforcement a program designed for tutorial makes available, investigations based on Science to study students and their learning processes. Prediction, clustering, mining data to discover associations, filtering data to enhance comprehension, and uncovering knowledge through the use of pattern were the five methodologies proposed by them. (Castro et al., 2007) proposed the following EDM subjects/tasks: program for assessing learner accomplishment, program for providing adjustment and counsel to scholars taking into consideration their individual differences in learning, methods of assessing teaching aids and academic programs online, programs designed to provide assessment reports in lessons delivered and received electronically to learners as well as teachers, and designing programs that will identify distinctive traits of students in acquisition of knowledge.

1) Forecasting

Forecasting focuses on building a prototype for making inference on a distinct portion of the data referred to as predicted variable, based on the integration of other portion of the data known as predictor variables. In prediction it required to have labels for the variable whose output is being determined for a restricted set of data. The label is a representation of information about the value of the output variable in distinct cases obtained from a reliable source. In some cases, it is important however in some cases to give consideration to the degree of reliability of the label's approximation.

In EDM there are two main uses of prediction. In some cases, studying the structures of a model is essential to prediction and it's a use of methods of prediction, which gives information concerning the fundamental construct. Research programs try to forecast the educational results of students and this is a shared approach in research programs (Romero et al, 2008). Another way by which methods of prediction are made use of is to forecast what the resulting value may be, in contexts where obtaining a label of that construct would not be desired. Considering a research that tried to examine the connection between the system of gaming and learning is an example of trying to thrive in an environment of interactive learning by utilizing characteristics of the system instead of learning the material. Review of the structure in twelve month of during which the learning program in various learning environment by a researcher, may not be amenable to a direct access, putting

to use methods that are not data-mining, whether a point in time a student is gaming (Baker et al., 2008) by the use of methods of observable label to label small sets of data, a prediction model was developed. This was by the use of data collected automatically from communications between the gaming software and the students for predictor variables. The problems the researchers seek to solve within the confines of whole group of raw facts and figures were made possible.

Regression, classification, and density estimation are mainly the three prediction types.

In classification, the predicted variable is a dual or discrete changeable value. Methods regarded as standard approaches are decision trees, logistic regression, and support vector machines. In regression, the forecasted changeable value is a changeable value that has varying quantities. Approaches for measuring the level of association widely employed in EDM regression that can be mapped onto a straight line, neural set of connections, and support vector machine regression. Regression on a straight line is the least difficult kind of regression where facts and figures are represented on a straight line. The value that assesses the level of association is derived through a method called least squares. A kind of regression that handles many forecaster variables (Multiple Regression) is the expanded form of linear regression and engages many changeable values. It is possible to transform nonlinear regression to linear regression. An example is polynomials.

The most widely used methods in determining the level of association between variables include polynomial regression, linear regression, multi-linear interpolation, robust regression, cascade correlation, radial basis functions (RBFs), multivariate adaptive regression splines (MARS), group methods data handling (GMDH), and regression trees. Methods of regression were employed for forecasting the degree of learners' comprehension based on the causes of their mistakes (Feng et al., 2005). In establishing comparative proficiency of sundry methods of teaching, (Feng et al., 2009) employed breakdown method of learning, method of knowledge discovery in data obtained from academic institutions, and methods of logistic regression for establishing the amount of knowledge acquisition that occurred as a result of dissimilar instructional technique being applied to the impartation of similar aptitude.

Linear correlation is a trendy approach used to evaluate a forecaster's appropriateness (Bradley, 2007). In general terms, measuring precision in percentages is not a desired choice in categorization. This is because precision estimates depend extremely on initial percent on hundreds of various groups. In the determination of a predictor's appropriateness, making sure that the various observations on a particular learner are dependent is not negotiable. Investigators in the field of EDM employ methods that go far beyond the ordinary for analysis to make up for incomplete dependence (Strube, 1985).

2) *Predicting Student Performance*

Estimating the value yet to be determined for a changeable entity that explains issues about the student is the essence of prediction. The predicted values typically include accomplishment, comprehension, and grade, which can either be statistical and suitable for regression or definite, which is suitable for classification. The purpose of regression is to verify the association that a dependent changeable entity has with a single or additional independent changeable entities (Draper and Smith, 1998). In the process of classification, entities are grouped on the basis of numerical data with regard to a singular or additional intrinsic features of the entity on the basis of a training set of entities that were formally categorized (Espejo et al., 2010).

Among the ancient and generally applied in EDM is student's accomplishment forecasting, in which many dissimilar methods as well as models are employed.

Methods of machine learning had been compared to carry out forecasting of accomplishment in a program of study, whether good or poor in ITS (Hamalainen and Vinni, 2006). Further contrasting was made of various step-by-step manner of mining data for the purpose of grouping learners, and forecasting their final scores in the context of data generated from using Moodle (Romero et al., 2008); for forecasting of learner accomplishment on the premise of attributes observed in user records (Minaei-bidgoli et al., 2003) and for predicting scholastic accomplishment of students in the university (Ibrahim and Rusli, 2007).

Using models of neural network of various kinds, the achievement of a learner at the end of a program has been forecasted (Gedeon and Turner, 1993); for the prediction of the amount of mistakes a learner will commit (Want and Mitrovic, 2002); for the prediction of accomplishment on the basis of marks obtained in quiz (Girones and Fernandez, 2006); to forecast scholars' grades, whether successful or not successful on the premise of logged data from Moodle (Delgado et al., 2006) and for the prediction of probable accomplishment of an applicant well-thought-out for entrance into an educational institution (Adebanjo et al., 2008). Bayesian networks were employed for the prediction of achievement of learners considered for enrollment (Haddawy et al., 2007); for modeling learner comprehension and forecasting of scholar accomplishment from a teaching scheme (Pardos et al., 2007); for predicting an upcoming graduate's CGPA using candidate's experience during enrollment (Hien and Haddawy, 2007); for modeling dual dissimilar methods to establish the possibility of correcting a question demanding a lot of aptitude (Pardos et al., 2008) and for predicting upcoming collective accomplishment in direct cooperative studies (Cooper et al., 2005); for forecasting accomplishment at the end of a period of studies in online system (Ayers and Junker, 2006); and for forecasting the result of an entity reaction (Desmarais et al., 2006).

Varied kinds of regulation-centered schemes were employed for forecasting learner a accomplishment in settings that acquire knowledge through electronic media (Nebot et al., 2006); to forecast student accomplishment on the basis of learning collections built (Chen et al., 2007); for forecasting, supervising and assessment of scholastic accomplishment of learners (Ogor, 2007); for predicting ratings at the end of a course of study on the premise of characteristics mined from logged data in online educational establishments (Shangping and Ping, 2008); for predicting learner accomplishment to offer well-timed tutorials in online e-learning environment (Chan, 2007); to forecast learner's score in web-based systems (Etchells et al., 2006).

Several regression techniques have been used to predict students' marks in an open university (using model trees, neural networks, linear regression, locally weighed linear regression and support vector machines) (Kotsiantis and Pintelas, 2005); for forecasting responsibility evaluation marks through linear regression projection patterns at end of the year (Anozie and Junker, 2006); to forecast learner accomplishment from logbook and quiz marks in online learning systems through multivariate regression pattern (Yu et al., 1999); for forecasting learner scholastic accomplishment by employing stepwise linear regression (Golding and Donalson, 2006); employing multivariate linear regression to forecast learning page duration (Arnold et al., 2005); multivariate regression used for detecting estimates capable of predicting good attainment that could predict success in institutions curricula (Martinez, 2001); regression and decision trees analysis were used to forecast learner contentment in higher institution (Thomas and Galambos, 2004); for forecasting examination results in distance learning programs (Myller et al., 2002); for forecasting when a student will get a question correct and association rules to guide a search process to find transfer models to predict a student's success (Freyberger et al., 2004); to forecast the likelihood of a learner providing the right response to question in ITS (Cetintas et al., 2009); for forecasting responsibility evaluation marks at year ending (Anozie and Junker, 2006), for predicting a learner's quiz mark (Feng et al., 2005) and for predicting the likelihood of a learner's subsequent answer to be right (Beck and Woolf, 2000).

Conclusively, correlation analysis was employed collectively to forecast accomplishment in web-based lessons (Wang and Newlin, 2002); to forecast a learner's final examination mark in web-based teaching (Pritchard and Warnakulasooriya, 2005) and to forecast the possibility of learners from secondary institutions excelling in university (McDonald, 2004).

3) *Relationship Mining*

Relationship mining, seeks to achieve the objective of determining the associations between variables, considering a set of data with a great number of variables. This could be by finding out the variables that associate strongly with a variable of specific interest, or by trying to know the strongest relationship between any two variables of interest.

Largely, relationship mining is of four types, namely mining to establish causal relationship (association rule), mining to detect the level of association (correlation mining), mining to identify sequence of patterns, and fundamental mining of data. Objectively, mining to detect a causal relationship is to obtain a rule of conditionality where the presence of a group of changeable values defines in general terms a unique value of another variable.

In mining to ascertain the degree of association, determining whether the bond is positive or negative, and whether the relationship between variables is linear is the prime purpose. In mining data to establish causal relationship, the objective is to determine if the occurrence of an experience accounts for another event occurring. This is achieved by examining the variation between both occurrences or observing the cause of one of the occurrences (Scheines et al., 1994).

Two decisive factors that relationships ought to comply with are: numerical consequence, and interestingness. Numerical assessments tools that are internationally accepted are often used for assessing numerical significance. As a result of the numerous assessments that are carried, there is always the need to regulate processes leading to the discovery of stochastic relationships. Post-hoc numerical techniques or modification can be used as a regulatory method for the number of assessments carried out. By the use of this technique there is high assurance that each relationship discovered was never probabilistic. Using the Monte Carlo method is another option for assessing the entire trend discovered to establish its probabilistic nature.

For the purpose of reducing the collection of conventions, the measure of degree of association, or underlying connections that the researcher has uncovered, it is essential to assess every discovery in the context of its relevance. A great number of noteworthy associations are contained in large volumes of data. The verification of the uniqueness of the findings, their proper representation of the data and the effort to reduce excessive related discoveries is referred to as interestingness. Among the distinct determinants of interestingness are coverage, cosine, conviction, confidence, lift, correlation, leverage, and support. Lift and cosine have been proposed by researchers to be uniquely applicable data from educational establishments (Merceron and Yacef, 2008).

4) *Discovery with Models*

Concerning the use of a model for the discovery of patterns, there is the building a model of an observable fact through clustering, forecasting, or in some instances development of knowledge. The model that is developed becomes a constituent in a further investigation, for instance prediction or mining to discover association.

In the case of prediction, in predicting a new variable, the predictions of the model developed serves as created model's predictions are used as variables that can be used for prediction. The representation of the constituents of

knowledge in a sphere have sequentially become the determinant for the assessments of student knowledge, which is by and large presented as a relation connecting tasks within the knowledge components and educational program.

Within the instance of mining to discover relationships, there is analysis of the relationships concerning further changeable elements and the projections of the model developed. A researcher can be enabled to examine the association involving a wide variety of observable constructs and a complex latent construct.

Frequently, employing models for discoveries influences the validity of a model's projection to be generalized from one context to another. Generalization depends on appropriate validation.

5) *Filtering of Raw Facts and Figures*

Filtering raw facts and figures is one of the fields of attraction for researchers in EDM. Individuals are capable of giving interpretations regarding raw facts and figures, in some instances, if properly represented, in ways outside the present capacity of absolutely automatic information extraction techniques. These processes are called information visualization methods in educational data mining. Nevertheless, the most widely used visualizations methods within educational data mining are often dissimilar to those typically applied in further ways of viewing knowledge challenges (Kay et al, 2006; Hershkovitz and Nachmias, 2008).

There are two main reasons for the filtering of raw facts and figures in EDM, namely categorization and for recognition. Information alternatively can be filtered from individuals' recognition, to assist in building and modeling future forecasting samples. In such an instance, sub-sections of a data set are representation of sub-sections of the set of data are done in text (visual format), and classification is done by coders. By and large, these labels then become the foundation for a predictor development. This method has a proof of accelerating improvements of predictive prototypes of complicated trends, is accelerated by this proven method (Baker and de Carvalho, 2008).

D. *Application of Educational Data Mining*

Although, using data mining in education is a recent research field, there are many works in this area. That is because of its potentials to educational institutes.

(Walters, and Soyibo, 2001) conducted a study to determine accomplishment of learners in secondary institutions on five Integrated Science progression proficiencies with accomplishment in connection to sex, rating, institute site, institute kind, kind of learner, and social and financial status. The results revealed that there was a positive significant relationship between scholastic accomplishment of the learner and the institution type.

(Hijazi and Naqvi, 2006) conducted a study on the learner accomplishment by selecting a sample of 300

comprising 225 males and 75 females, a collection of college students. The hypothesis was stated as Learner's posture towards attending lesson, time used to learn each day, family earning, age of learner's mother and learner's mother's educational status are considerably connected with learner accomplishment. By means of simple linear regression analysis, it was discovered that elements like learner's mother's educational status and learner's family earnings were highly connected with learners' scholastic accomplishment.

(Bharadwaj and Pal., 2011) undertook a study in which Bayesian categorization was used to assess learner accomplishment. 300 learners were selected as sample of which 226 were males and 74 were females. A total number of 17 characteristics were considered to ascertain which of them have impact on the learner accomplishment in Bachelor of Computer Application program in an Indian higher institution of learning called Dr. R. M. L. Awadh University. Questionnaire and institution's database were used to gather statistics on learner's scholastic, personal and social and financial characteristics. Learner's scores were gathered from the institution's assessment unit. Learner's rating in senior secondary exam, residence, channel of receiving tutorial, prerequisite of mother, other addiction of learner, family yearly earning and status of learner's family were discovered to be extremely linked with scholastic accomplishment of the learner. The study concluded that scholastic accomplishment of learners does not always depend on the learner's own endeavor.

(Tair and El-Halees, 2012) conducted a case study in educational data mining. Their study showed how data mining can be used in higher education particularly to improve graduate students' performance. They used graduate students' data collected from the college of Science and Technology in Khanyounis. The data covered fifteen years period. They applied data mining techniques to discover knowledge. Particularly, they discovered association rules and sorted the rules. Then they used two classification methods which are Rule Induction and Naive Bayesian classifier to forecast the status of the graduate student. Also they clustered the students using K-Means clustering algorithm. Finally, they used outlier detection to identify all outliers in the data.

(Kotsiantis et al., 2004) applied several step-by-step methods of DM for predicting the accomplishment of computer science students in distance learning course in a higher institution. For every individual learner, a number of personal characteristics e.g. Sex, age, marital status and accomplishment qualities such as scores were used as parameters of a dual classification of pass or fail. The finest result was achieved by a Naive Bayes method which gave precision of 74%. In addition, discovery was made that grades obtained in former institute of scholastic accomplishment has much greater influence than personal characteristics.

(Pardos et al., 2006) gathered facts and figures from a web-based system of teaching in Math quiz. Regression

analysis was used and the objective was to forecast the marks on the basis of personal aptitude of the learner. Bayesian Networks were employed by the researchers to obtain the finest outcome of 15% as a forecast mistake.

(Ayesha et al. 2010) employed a method of data mining called step-by-step k-means clustering to forecast behaviors of learners in a learners' record comprising class tests, middle of term and end of term exams and projects. This connected statistics would be transferred to the class instructor for the organization of final exam. The investigation helped the instructors to bring down the declining percentage by applying the right actions at appropriate time to enhance learner accomplishment.

(Moriana et al. 2006) investigated the likely impact of extra-curricular activities like study-related and/or sports-related activities on the scholastic accomplishment of learners in a Spanish secondary school. The sample selected comprised 222 learners. The learners were picked out of 12 distinct institutions. Classification of the sample was done in two categories based on what students engaged in when school was not in session. The data was analyzed using Analysis of variance (ANOVA) to substantiate the impact of activities that are non-academic on the scholastic accomplishment. Observation showed that the category of students engaged in activities that do not relate to academic work had higher scholastic accomplishment.

(Baradwaj and Pal, 2011) employed data mining technique known as categorization, and specifically decision tree approach to evaluate students' scholastic accomplishment. The purpose of their study was to mine knowledge that throws light on learner's scholastic accomplishment in exam at semester ending. They used students' data from the student database including attendance, class test, seminar, and assignment scores. This study helped earlier in identifying the dropouts and students who need special attention and allows the teacher to provide appropriate counsel and remedy.

(Al-Radaideh et al. 2006) applied data mining techniques, particularly classification to help in improving the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. The extracted classification rules are based on the decision tree as a classification method; the extracted classification rules are studied and evaluated. It allows students to predict the final grade in a course under study.

An in-depth investigation was undertaken by employing EDM technique for assessing characteristics of students in acquisition of knowledge. The research was to establish the usefulness of mining data in tertiary institutions to enhance accomplishment of students. Students' data was collected consisting of demographic and scholastic statistics, and data related to the subject of study from electronic learning system. The researcher utilized data methods of data extraction (decision tree) for the discovery of numerous varieties of information, including rules for association and

rules for classification. The researcher further categorized the students into clusters for the purpose of analyzing outliers. Eventually, the researcher displayed the way by which one can take advantage of the information uncovered for enhancement of learners' accomplishment (El-Halees, 2008).

E. Visualization of Data

Data visualization has the purpose of visualization of putting indispensable information at the center of attention and support in taking of decisions. In institutes of learning, tutor and program supervisors are capable of analyzing how students behave toward their program of study. They are also capable of using information from such analysis to make inference about how students learn. Visualization of information and statistics have been the two methods generally used for performing such tasks.

Statistics is a scientific discipline mathematically inclined. It involves the gathering, examination, reading, and display of facts and figures (Freedman et al., 2007). It is comparatively simple to basically analyze and interpret facts and a program like SPSS, used for statistical analysis of data. The complete features of data can be presented in abstracts and descriptions of learner behaviors when data generated from academic settings are analyzed graphically (Wu et al., 2002). Tutors therefore have preference for data analysis on facts and figures on teaching activities which are not difficult to explain (Zinn and Scheuer, 2006).

F. Summary

Mining of Data involves the use of categories of tools and techniques to uncover unrecognized samples and associations found in the data (Dunham, 2003). DM similarly is by and large a single stage in the process of discovering information, in which institutions desire to be well informed in taking decision using fresh data. It can therefore be said that in taking major decisions, DM as well as discovery of information plays a crucial and inevitable role. Considering the import of the definition and explanation of data mining by the various authors, all of them seemed to be dwelling on the same idea. The pivotal issue in all definitions is the discovery of valuable information from patterns, trends, and relationships extracted from huge or complex data sets with a primary purpose of making informed decisions that will improve efficiency.

Educational data mining (also referred to as "EDM") is defined as the area of scientific inquiry centered on the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to understand students and the settings which they learn better.

Forecasting focuses on building a prototype for making inference on a distinct portion of the data referred to as predicted variable, based on the integration of other portion of the data known as predictor variables. In prediction it required to have labels for the variable whose output is being determined for a restricted set of data. The

label is a representation of information about the value of the output variable in distinct cases obtained from a reliable source. In some cases, it is important however in some cases to give consideration to the degree of reliability of the label's approximation.

Several regression techniques have been used to predict students' marks in an open university (using model trees, neural networks, linear regression, locally weighed linear regression and support vector machines) (Kotsiantis and Pintelas, 2005); for forecasting responsibility evaluation marks through linear regression projection patterns at end of the year (Anozie and Junker, 2006); to forecast learner accomplishment from logbook and quiz marks in online learning systems through multivariate regression pattern (Yu et al., 1999); for forecasting learner scholastic accomplishment by employing stepwise linear regression (Golding and Donalson, 2006); employing multivariate linear regression to forecast learning page duration (Arnold et al., 2005); multivariate regression used for detecting estimates capable of predicting good attainment that could predict success in institutions curricula (Martinez, 2001); regression and decision trees analysis were used to forecast learner contentment in higher institution (Thomas and Galambos, 2004); for forecasting examination results in distance learning programs (Myller et al., 2002); for forecasting when a student will get a question correct and association rules to guide a search process to find transfer models to predict a student's success (Freyberger et al., 2004); to forecast the likelihood of a learner providing the right response to question in ITS (Cetintas et al., 2009); for forecasting responsibility evaluation marks at year ending (Anozie and Junker, 2006), for predicting a learner's quiz mark (Feng et al., 2005) and for predicting the likelihood of a learner's subsequent answer to be right (Beck and Woolf, 2000).

Conclusively, correlation analysis was employed collectively to forecast accomplishment in web-based lessons (Wang and Newlin, 2002); to forecast a learner's final examination mark in web-based teaching (Pritchard and Warnakulasooriya, 2005) and to forecast the possibility of learners from secondary institutions excelling in university (McDonald, 2004). It was on this premise that the methodology was framed.

III. METHODOLOGY

It is widely accepted that Research methodology involves how to identify the outcome of a particular problem of an issue under study. Researchers apply various diverse measures in tackling research problems of varying sources. Therefore, how a solution is obtained or the process of searching for a research problem is termed Methodology (Industrial Research Institute, 2010).

The methodology by which a research is undertaken is orderly and organized and it is a discipline concerned about ways a research is undertaken. Importantly, how a researcher describes the work, explains and predicts trends

applying scientific and logical methods to acquire information from the data is termed a research methodology.

It is evident that research methods, an integral portion of the research methodology comes in different aspects. Therefore, when research methodology is mentioned it involves not only research methods but also considers the reason behind the techniques applied so as to enable the researcher or anyone evaluate the decisively the outcomes of the research. The reason for a research to be carried out, the definition of the problem, formulation of hypothesis, the data gathered and the process used, the reason behind the evaluation methods adopted and other issues of concern are generally answered once studying a problem under research applying research methodology.

Researchers when employing Research Methodology, orderly identify particular questions and find solutions to them until they arrive at a conclusion. When the research done is not sequential it reduces the possibility of obtaining a meaningful outcome (Industrial Research Institute, 2010).

A. Research methods

All the different processes, structures and calculations employed by researchers in studying a research are termed as research methods.

Generally, its structured, technical and logical which consists of theoretic processes, investigational studies, numerical methods, etc. These aid researchers in gathering information to enable them attain a suitable solution. Mostly, technical methods of researching require gathered evidences, calculations and observing of events and necessarily human judgment. Hence it is adopted if it is proven by experimenting.

It is widely accepted that Research methodology involves how to identify the outcome of a particular problem of an issue under study. Researchers apply various diverse measures in tackling research problems of varying sources. Simply put, the entirety of processes and techniques utilized by researcher in their course of work are research methods.

The main object of carrying out researches is to arrive at an answer; hence the information gathered as well as the undetermined parameters of problem at hand should co-ordinate, to increase the possibility of achieving a solution. With respect to this, there are 3 categories for grouping methods of research:

- i. Methods associated with collecting information.
- ii. Techniques that relate the available information with the undetermined factors, usually called statistical methods.
- iii. Evaluating Methods employed to analyze how accurate the results are.

B. Research Paradigms

Traditionally, approaches to research are divided into two main paradigms, namely, the Qualitative/Interpretivist and the Quantitative/positivist approach.

Essentially, the worldwide view accepted by researching happenings is referred to as a research paradigm (Creswell, 2009). Generally, research paradigm is a viewpoint of study which involves major themes, methods of looking for responses, and fundamental suppositions underlying the research methods.

As said by (Johnson and Christensen, 2005), research paradigm is a standpoint founded by a group of agreed conventions, standards, ideas and practice. Simply put, the way in which researchers perceive or understand growing knowledge is referred to as a paradigm. Importantly, research paradigms aid researchers in conducting studies effectively.

Combining methods of researching as well as philosophies in research and applying the right research paradigm and philosophy assist researchers get rid of factor that impedes the effective implementation of techniques such as finance, restrictions among others. Three distinct terms are mentioned in reference to research paradigm they are: positivism, interpretivism and realism research philosophies (Mcnabb, 2007).

Positivism Philosophy: This specifically designed on the extremely organized method that makes generalizing and quantifying observations possible as well as in analyzing results aided by some organized data processes. This is generally applied in natural science making it a method objectively and analytically based. This process involves natural science philosophies which include constant philosophy, general laws with the perspective of all naturally occurring things (Saunders, 2003).

Every statistics concerning the research under study can be gathered by researchers with the aid of positivism philosophy. The function of researchers in the philosophy of research is of essence to the research study as proposed by (Saunders, 2003) as the researchers offer an important task of objectively analyzing the gathered information to produce positive outcomes as well as accomplish the goal and objective of the research.

Interpretive Philosophy: It of the belief that the societal domain of managing and commerce is sophisticated in formulating theories and laws for example in the natural science. The aspect of positivism philosophy that deals with extensive thinking is interpretive philosophy. It is of the view that, there exist numerous suitable solutions to every research issue from of a simple statistics (Johnson and Christensen, 2010).

Interpretive Philosophy is of a vital function in obtaining out comes from information gathered. With Interpretive Philosophy researchers do not necessarily interrelate with their surroundings but try to understand by making judgments and making sense out of them. An individual's character can be influenced by difference in the standards of living, different societal and traditional settings, character differences and family inclination etc. (Saunders, 2003).

Realism Philosophy: it is as well a significant philosophy structured on human interrelation, ethics and opinions. Realism Philosophy is concentrated on the views present in a location. In realism philosophy, it is accepted that societies' way of making judgments as well as their conduct, is influenced by the impartial reality that exist and that in natural science humans are not the matter of investigation. It can therefore be concluded that realism philosophy reveals the way humans respond to realistic scenarios (Johnson and Christensen, 2010).

C. Quantitative Research/ Positivism

Quantitative in a wider perspective comprises of gathering data and illustrating how theories relate to researches taking it from an objective and realistic point of view. Quantifying information can be said to be the objective of quantitative research. Results can be generalized in quantitative research as to view a sampled population as the whole and measuring ideas and views sampled. Against this backdrop, this thesis therefore adopts the Quantitative/Positivism research paradigm in which the researcher made use of quantitative research approach by applying statistical analysis to numerical data (students' marks) that resulted in various parameters which were interpreted and generalized for the entire year group. This research paradigm is considered the most appropriate because the study is based on the measurement of quantity or amount. The research process is expressed in terms of one or more quantities, and hence the result of this research is a set of numbers.

D. Research Design

A research design represents a plan, structure, and strategy of investigation conceived so as to obtain answers to research questions and to control variance. In order to achieve the primary objective of the study, which is to predict the performance of students in WASSCE, a cross sectional study which is a form of a Quantitative research was employed for this study. Mawuli School was selected as the area for the study.

A cross sectional survey approach provides a quantitative or numeric description of trends, attitudes, or opinions or a population by studying a sample of that population. It also deals with comprehensive enquiry into part of a subject with similar attributes to the entire subject under study to obtain an understanding of the whole subject.

To obtain the required data for the study, the researcher was allowed to have direct access to students' academic records upon request from the Academic Affairs Office of Mawuli School after having explained and assured the head that the records would not be used for any other purpose apart from academic purposes.

Quantitative analyses was done on the main themes and supported with statistical presentation of actual results. The major findings were also summarized in line with objectives of the study and recommendations drawn.

1) Population and Sample Size

All second year students of Mawuli SHS formed the target population for the study. In line with the objectives of the study, information bordering on the various variables of interest (students' scores/grades in BECE, students' scores/grades for the first and second years in SHS) was gathered on the students. A sample size of 180 was chosen from a population of 700 based on $\pm 7\%$ level of precision and 95% confidence level. Ten sets of data were considered; first and second year marks of students in Mathematics, English Language, Integrated Science, Social Studies, and average scores in core subjects in BECE and SHS. The data was gathered from the 2012 to 2015-year group.

2) Sampling Procedure

Each of the seven programs offered in Mawuli School form a cluster which is homogeneous in terms of the core subjects studied under each. A probability sampling technique, precisely simple random sampling was used to select records from each cluster. This sampling technique was deemed appropriate because the sampling units could be identified. In other words, there is a finite number of students in each program and a sampling frame could thus be formed. Furthermore, every student has an equal chance of being selected for the sample, and hence the use of stochastic technique.

3) Preparing Data for Analysis

Data cleaning was carried out on the data collected. This refers to the preprocessing of data in order to treat *missing values* (e.g., by replacing a missing value with the most commonly occurring value for that attribute, or with the most probable value based on statistics).

This pre-processing stage is the preparation of data before the application of data mining tool and it consists on a set of operations, which goal is to improve data quality and data mining methods. This stage is really important for successful data mining, as real world data is "dirty", i.e., it mainly consists of incomplete (e.g. missing attributes), inconsistent (e.g. different naming) and noisy data (e.g. data with errors). In order to summarize the data, students' marks in BECE and SHS were converted to averages. Only the core subjects were considered in this case.

4) Data Resources

The data resources can be classified as secondary or primary (Dawson 2002). The empirical data for this study was gathered from a primary source using a probability approach. The researcher collected continuous data comprising marks obtained by students in first and second years of their study.

5) Data Analysis Procedure

"Statistics is a mathematical science concerning the collection, analysis, interpretation or explanation, and presentation of data" (Freedman et al., 2007).

The data collected for the research was analyzed using correlation analysis and linear regression analysis. SPSS v16 was used for analysis of the data.

Descriptive statistics was used to give numerical and graphic procedures to summarize the data in clear and understandable way. Using the numerical approach, statistics such as the mean and standard deviation were computed for all the four core subjects. Also, standard deviations for the predicted mean marks were computed. These statistics convey information about the average. The standard deviation was computed to explain the extent to which the mean has been influenced profoundly by one or more extreme data points (referred to as an outlier). It was to show the relation that set of scores has to the mean of the sample. It was calculated as the square root of the sum of the squared deviations from the mean divided by the number of scores minus one.

Other statistics computed include the coefficient of determination, coefficient of regression, minimum and the maximum predicted values for each core subject.

The coefficient of determination (R-square) was used to evaluate the model fit of the regression equation. That is, how good all the independent variables are at predicting the dependent variable. The value of R-square ranges from 0.0 to 1.0 and can be multiplied by 100 to obtain a percentage of variance explained, and this was carried out as such in this research.

The coefficients from the regression equations represent the strength and direction of the relationship between the independent and dependent variables.

Graphical method also was used for identifying patterns in the data. Histogram was used to show the distribution of students' marks in the various subjects under consideration.

The interpretation of relationship between the variables (BECE grades and SHS grades) were based on the premise that the closer the data points come when plotted to making a straight line, the higher the correlation between the two variables, or the stronger the relationship. If the data points make a straight line going from the origin out to high x- and y-values, then the variables are said to have a positive correlation. If the line goes from a high-value on the y-axis down to a high-value on the x-axis, the variables have a negative correlation.

Data collected on students' academic performance in all core subjects was summarized into average marks. Two sets of such average scores, one set for first year and another set for second year were used as the independent and dependent variables respectively in the regression equation to build a model that would predict students' final mark/grade.

6) *Regression Analysis*

Regression was used to fit an equation to the dataset. This data mining technique/statistical tool was used for the investigation of relationships between the variables. It was to establish a causal relationship between the dependent or outcome variable (final scores) and the predictors (second

year scores). This means the estimation or prediction of the unknown value of one variable (final grade) from the known value of the other variable (second year scores).

7) *Simple Linear Regression Analysis*

The simplest form of regression, linear regression, uses the formula of a straight line ($y = ax + b$) and determines the appropriate values for a and b to predict the value of y based upon a given value of x . The equation that describes how y is related to x is known as the regression model. Simple linear regression was used to predict values of one variable, given values of another variable. The assumption that underpinned the use of this linear regression is that the target attribute is a linear function of other, mutually independent attributes. However, the model is very flexible and can work well, even if the actual dependency is only approximately linear or the other attributes are weakly correlated.

In this research, simple linear regression was used to predict the marks of students in their final exam given the numerical data of their first year and second year examination results. In this case the first year and second year marks of students were the independent variables and the dependent variable was the final exam mark of students.

8) *Regression Equation*

Suppose we have a sample of size ' n ' and it has two sets of measures, denoted by x and y . We can predict the values of ' y ' given the values of ' x ' by using the equation, called the regression equation, $Y = aX + b$ where the coefficients a and b are given by

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{\sum y - b \sum x}{n}$$

For the purpose of this research x is the marks obtained by students in first year, y is the marks obtained by students in second year, n is the sample size (180) and Y is the predicted mark. A prediction model was then formed for the four core subjects studied by students. The computation that went into the formation of each model is presented as follows:

9) *Prediction Model for Students' Final Marks in Mathematics*

The y-intercept is given as 16.278 and the Slope is given as 0.532 in table 4.2.

The slope and the y-intercept were substituted in the following linear equation to predict students' final score in Mathematics: $Y = aX + b$. In this case the values of a , b , x , and y were as follows:

$$a = 0.532$$

$$b = 16.278$$

$$X = \text{First Year Score}$$

$$Y = \text{Second Year Score}$$

To predict students' final scores, the values of a , and b were substituted into the following linear equation: $Y = aX + b$ and the resulting equation is:

$Second\ Year\ Score = a * First\ Year\ Score + b$ and the regression equation is given by
 $Final\ Year\ Score = 0.532 * Second\ Year\ Score + 16.278$, where

$$\begin{aligned} a &= 0.532 \\ b &= 16.278 \\ X &= Second\ Year\ Score \\ Y &= Final\ Score \end{aligned}$$

The regression equation $Final\ Year\ Score = 0.532 * Second\ Year\ Score + 16.278$ was used to project students' final WASSCE scores in Mathematics.

10) Prediction Model for Students' Final Marks in English Language

The y-intercept is given as 30.267 and the slope is given as 0.465 in table 4.5. The slope and the y-intercept were substituted in the following linear equation to predict students' final score in English Language: $Y = aX + b$. in this case the values of a , b , x , and y were as follows:

$$\begin{aligned} a &= 0.465 \\ b &= 30.267 \\ X &= First\ Year\ Score \\ Y &= Second\ Year\ Score \end{aligned}$$

To predict students' final scores, the values of a , and b were substituted into the following linear equation: $Y = aX + b$

This gives the equation: $Second\ Year\ Score = a * First\ Year\ Score + b$ and the resulting regression equation given by $Final\ Year\ Score = 0.465 * Second\ Year\ Score + 30.267$, where

$$\begin{aligned} a &= 0.465 \\ b &= 30.267 \\ X &= Second\ Year\ Score \\ Y &= Final\ Score \end{aligned}$$

The regression equation $Final\ Year\ Score = 0.465 * Second\ Year\ Score + 30.267$ was used to project students' final WASSCE scores in English Language.

11) Prediction Model for Students' Final Marks in Integrated Science

The y-intercept is given as 2.445 and the slope is given as 0.865 in table 4.5. The slope and the y-intercept were substituted in the following linear equation to predict students' final score in English Language: $Y = aX + b$. in this case the values of a , b , x , and y will be as follows:

$$\begin{aligned} a &= 0.865 \\ b &= 2.445 \\ X &= First\ Year\ Score \\ Y &= Second\ Year\ Score \end{aligned}$$

To predict students' final scores, the values of a , and b were substituted into the following linear equation: $Y = aX + b$. This gives the equation $Second\ Year\ Score = a * First\ Year\ Score + b$ and the resulting regression equation given by $Final\ Year\ Score = 0.865 * Second\ Year\ Score + 2.445$, where

$$\begin{aligned} a &= 0.865 \\ b &= 2.445 \\ X &= Second\ Year\ Score \\ Y &= Final\ Score \end{aligned}$$

The regression equation $Final\ Year\ Score = 0.865 * Second\ Year\ Score + 2.445$ was used to project students' final WASSCE scores in Integrated Science.

12) Prediction Model for Students' Final Marks in Social Studies

The y-intercept is given as 24.604 and the slope is given as 0.684 in table 4.11. The slope and the y-intercept were substituted in the following linear equation to predict students' final score in Social Studies: $Y = aX + b$. in this case the values of a , b , x , and y were as follows:

$$\begin{aligned} a &= 0.684 \\ b &= 24.604 \\ X &= First\ Year\ Score \\ Y &= Second\ Year\ Score \end{aligned}$$

To predict students' final scores, the values of a , and b were substituted into the following linear equation: $Y = aX + b$. This gives the equation $Second\ Year\ Score = a * First\ Year\ Score + b$ and the resulting regression equation given by $Final\ Year\ Score = 0.684 * Second\ Year\ Score + 24.604$, where

$$\begin{aligned} a &= 0.684 \\ b &= 24.604 \\ X &= Second\ Year\ Score \\ Y &= Final\ Score \end{aligned}$$

The regression equation $Final\ Year\ Score = 0.684 * Second\ Year\ Score + 24.604$ was used to project students' final WASSCE scores in Social Studies.

The computation of the coefficient of b is explained by the algorithm below:

1. Sum the values of x
2. Sum the values of y
3. Multiply the values of x and y and find the sum
4. Multiply the result in 3 by the sample size n
5. Multiply the result in 1 by the result in 2
6. Find the squares of the values of x and sum them
7. Multiply the result in 6 by the sample size n
8. Find the square of the result in 1
9. Subtract the result in 5 from the result in 4
10. Subtract the result in 8 from the result in 6
11. Divide the result in 9 by the result in 10

The computation of the coefficient of a , is explained in the following algorithm:

1. Sum the values of y
2. Sum the values of x and multiply the result by b
3. Subtract the result in 2 from the result in 1
4. Divide the result in 3 by the sample size n
5. The symbol Y refers to the predicted value of y from a given value of x from the regression equation.

13) Assumptions

In order to conduct a linear regression analysis, the following assumptions about the data were made:

1. Linearity: It is assumed that the relationship between the independent and dependent variables is linear.
2. Normality: It is assumed that the residuals of the variables are normally distributed. That is, the errors in the prediction of the value of Y (the dependent variable) are distributed in a way that approaches the normal curve.
3. Independence: It is assumed that the errors in the prediction of the value of Y are all independent of one another (not correlated).
4. Homoscedasticity: It is assumed that the variance around the regression line is the same for all values of the independent variables.

14) Correlation Analysis

The relationship between the marks obtained by students in Basic Education Certificate Examination (BECE), and the marks of students at the Senior High School (SHS) level was studied using correlation analysis. Correlation as a tool was used to study and measure the extent of the relationship between the two variables.

15. Coefficient of Correlation

One of the statistics used in the analysis of students' data is the coefficient of correlation, ' r ', which measures the degree of association between the two values of related variables given in the data set. That is the degree of association between the marks obtained by students in BECE, and the marks of students at the SHS. It takes values from $+1$ to -1 . If two sets of data have $r = +1$, they are said to be perfectly correlated positively; if $r = -1$ they are said to be perfectly correlated negatively; and if $r = 0$ they are uncorrelated.

The coefficient of correlation, ' r ', is given by the formula

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$

In the formula above, n is the sample size (180), x is the marks obtained by students in BECE, y is the marks obtained by students in SHS, and r is the coefficient that determines the degree of association between students' performance in BECE and SHS.

The following algorithm explains how the coefficient (r) is calculated:

1. Multiply the values of x and y and sum them
2. Sum the values of x
3. Sum the values of y

4. Multiply the result in 2 (sum of the values of x) by the result in 3 (sum of the values of y)
5. Find the squares of the values of x and sum them
6. Find the square of the result in 2 (sum of the values of x)
7. Find the squares of the values of y and sum them
8. Find the squares of the result in 3 (sum of the values of y)
9. Multiply the result in 1 by the sample size n , and subtract the result in 4
10. Multiply the result in 5 (sum of the squares of x) by the sample size n , and subtract the result in 6 (sum of the squares of x)
11. Find the square root of the result in 10
12. Multiply the result in 7 (sum of the squares of y) by the sample size n , and subtract the result in 8
13. Multiply the result in 11 by the result in 12
14. Divide the result in 9 by the result in 13

E. Summary

Research methodology is a way to systematically solve a research problem. It may be understood as a science of studying how research is done scientifically. In it we study the various steps that are generally adopted by a researcher in studying his research problem along with the logic behind them.

All the methods used by a researcher during a research study are termed as research methods. They are essentially planned, scientific and value-neutral. They include theoretical procedures, experimental studies, numerical schemes, statistical approaches, etc. Research methods help us collect samples data and find a solution to a problem.

According to (Johnson and Christensen, 2005), research paradigm is a perspective that is based on the set of shared assumptions, values, concepts and practices. In other world, paradigm can be defined as a function of how researcher thinks about the development of knowledge. Research paradigm is a combination of two ideas that are related to the nature of world and the function of researcher. It helps researcher to conduct the study in an effective manner. Traditionally, approaches to research are divided into two main paradigms, namely, the Qualitative/Interpretivist and the Quantitative/positivist approach. This thesis therefore adopts the Quantitative/Positivism research paradigm in which the researcher made use of quantitative research approach by applying statistical analysis to numerical data (students' marks) that resulted in various parameters which were interpreted and generalized for the entire year group. This research paradigm is considered the most appropriate because the study is based on the measurement of quantity or amount. The research process is expressed in terms of one or more quantities, and hence the result of this research is a set of numbers.

A research design represents a plan, structure, and strategy of investigation conceived so as to obtain answers to research questions and to control variance. In order to achieve the primary objective of the study, which is to predict the performance of students in WASSCE, a cross

sectional study which is a form of a Quantitative research was employed for this study. Mawuli School was selected as the area for the study.

In conclusion the research design and various methods used to collect and analyze the numerical data produced the intended results which were discussed in the next chapter.

IV. DATA ANALYSIS AND RESULTS

The data was analyzed and the results presented in formulas, tables and charts. Various interpretations of the result follow every table and chart. Each result was examined in terms of the objectives and the research questions of the study.

A. Regression Analysis of Students' Marks in Mathematics

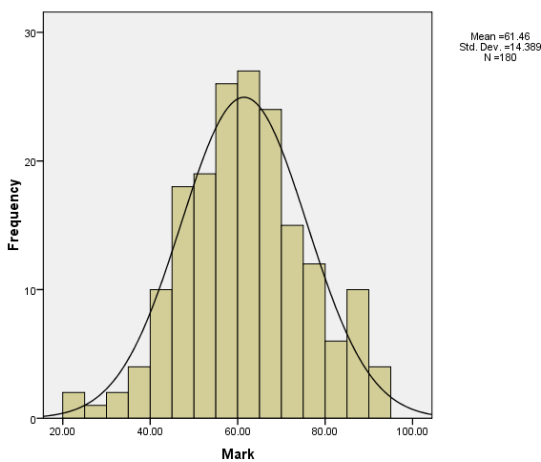


Figure 4.2. Distribution of Marks Obtained in Mathematics by Students in First Year

The mean of the distribution is 61.46% and the standard deviation is 14.39. Even though the mean is 61.46%, there is a significant number of students that fell below 61.46% (figure 4.2). The standard deviation also shows a wide variation of individual marks from the mean.

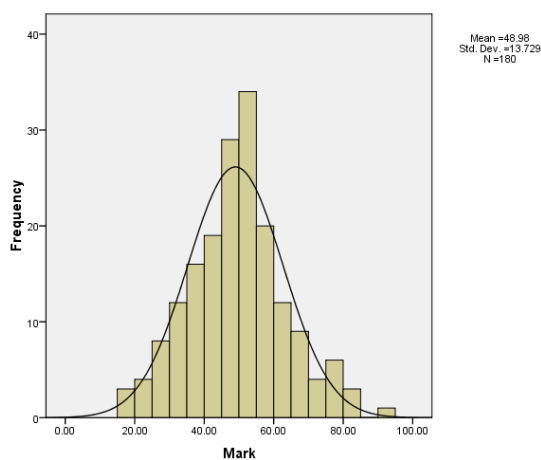


Figure 4.3. Distribution of Marks Obtained in Mathematics by Students in Second Year

The mean of the distribution is 48.98% and the standard deviation is 13.73. Students' performance in Mathematics declined considerably in second year giving a mean value of 48.98%, and a good number of students fell below the pass mark of 44%. Also the standard deviation is quite high, indicating the wide dispersion of individual marks from the mean.

The following tables present results of simple regression. R Square (.311) indicates that this model accounts for 31% of total variation in the data; the proportion of the variation that is explained by the model. R (0.558) is the absolute value of correlation coefficient. Adjusted R Square is the value adjusted for the number of variables in the regression model.

TABLE 4.1 MODEL SUMMARY FOR MATHEMATICS

Model Summary

Table 4.1 Model Summary for Mathematics

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.558	.311	.307	11.42813

a. Predictors: (Constant), First Year Scores

From the model R indicates a moderate correlation between the dependent and independent variable. This means there is a moderate degree of association between the marks obtained by students in second year and the marks they obtained in first year. 31.1% of the variance in second year scores of students in Mathematics is explained by the independent variable (first year scores of students in Mathematics). What this means is that aside other factors that affect performance, 31.1% of the effect is associated with students' performance in the first year, and this percentage is very significant to be considered.

TABLE 4.2. COEFFICIENTS FOR MATHEMATICS

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	16.278	3.747		4.345	.000
	First Year Scores	.532	.059	.558	8.962	.000

a. Dependent Variable: Second Year Score

The t-statistic presented in table 4.2 is 4.345 with a significance level less than 0.0001 (p<0.0001). The farther away this value is from zero, the more likely it is that the effect is statistically significant. This implies that the effect of the first year scores on the second year scores is not likely due to random chance, and that there is underlying repeatable cause. The conclusion is that students' second

year performance in Mathematics depended on their performance in the first year.

TABLE 4.3. RESIDUAL STATISTICS FOR MATHEMATICS

Residuals Statistics					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	26.9187	65.7578	48.9778	7.65554	180
Residual	-3.64567E1	27.24223	.00000	11.39617	180
Std. Predicted Value	-2.881	2.192	.000	1.000	180
Std. Residual	-3.190	2.384	.000	.997	180

a. Dependent Variable: second year score

From table 4.3, the minimum predicted value of 26.9% indicates that some students fell below the pass mark of 44. This situation is very bad in the sense that there is too much gap between this value and the pass mark and with the standard deviation 7.65554, the probability of many students failing is high. This prediction reflects students' second year performance which recorded 35% fail, and about 52.8% falling below the average mark of 49%.

The maximum predicted value is 65.7%, which can be graded as C4 by the West African Examination Council (WAEC) scheme for West African Senior Secondary Certificate Examination (WASSCE). The mean is predicted as 48.9%, which comparatively is almost the same as the mean for the distribution of students mark in the second year. It is however lower than the mean for the distribution of students marks in the first year (mean = 61.5%), and it is an indication of a serious fall in performance. It can therefore be concluded that students are not performing high in Mathematics and there is indication of further decline as predicted by the model.

B. Regression Analysis of Students' Marks in English Language

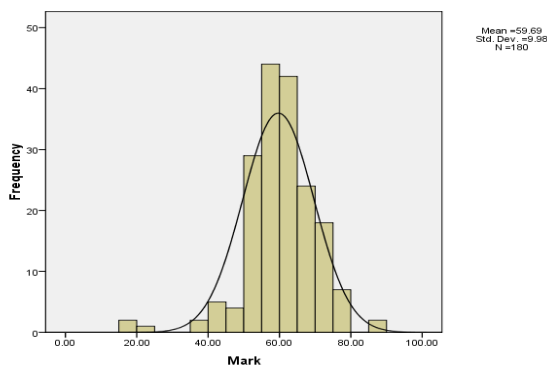


Figure 4.5. Distribution of marks obtained in English Language by students in first year

The mean mark is 59.69% and the standard deviation is 9.98. Some few students could not obtain the pass mark of 44 as shown on the chart. This represents 6.1%, and more so about 59% of students fell below the average.

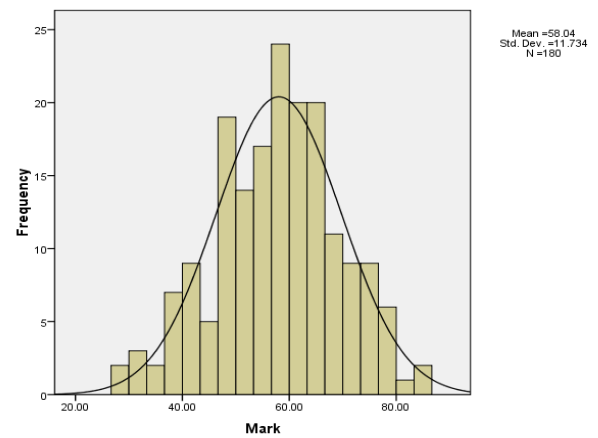


Figure 4.6. Distribution of marks obtained in English Language by students in second year

The mean of the distribution is 58.04% and the standard deviation is 11.734. There is a decline in performance over that of first year as seen on the chart, given the pass mark of 44% as we see more students falling below the pass mark.

TABLE 4.4 MODEL SUMMARY FOR ENGLISH LANGUAGE

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.396	.157	.152	10.80657

a. Predictors: (Constant), First Year Mark

From the model summary, R is 0.396, an indication of a low correlation between the dependent and independent variable. This model accounts for about 15% of total variation in the data, (adjusted R Square is 0.152). This means first year marks of students account for 15.2% of variation in second year marks of students in English Language. Standard error of the estimate is about 10.8.

TABLE 4.5. COEFFICIENTS FOR ENGLISH LANGUAGE
Coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
1 (Constant)	30.267	4.898		6.180	.000
Firstyear-mark	.465	.081	.396	5.749	.000

a. Dependent Variable: Second Year Mark

Form table 4.5, t-value is 6.180. The level of significance is less than 0.0001, an indication that the effect of students' first year performance on their second year performance is not due to likely random chance, and is statistically significant.

Table 4.6 Residual Statistics for English Language

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	37.7116	70.7460	58.0389	4.64345	180
Residual	-2.48573E1	25.28843	.00000	10.77634	180
Std. Predicted Value	-4.378	2.737	.000	1.000	180
Std. Residual	-2.300	2.340	.000	.997	180

a. Dependent Variable: Second Year Mark

Residuals statistics show that the predicted minimum, maximum, and mean values are 37.3%, 70.7%, and 58.03% respectively. The standard deviation for the predicted values is 4.6. Comparing the predicted mean with the mean of the distribution of students marks in both first and second year (figure 4.5 and 4.6), it can be seen that they are almost the same. This statistic reveals the reliability of the regression model. Given the minimum predicted value of 37.3% is indicative of the propensity of some students failing in English Language. In fact, this reflects students' performance in second year which recorded 13.3% failure (F9), about 52.8% falling below the average mark of 58.04%. The maximum predicted value can be converted to a grade B3 according to WAEC grading scheme for WASSCE. This reflects students' second year performance which recorded much decline. This model indicates a better performance in English Language when compared with Mathematics pointing to more work that needs to be done in Mathematics to remedy the situation.

C. Regression Analysis of Students' Marks in Integrated Science

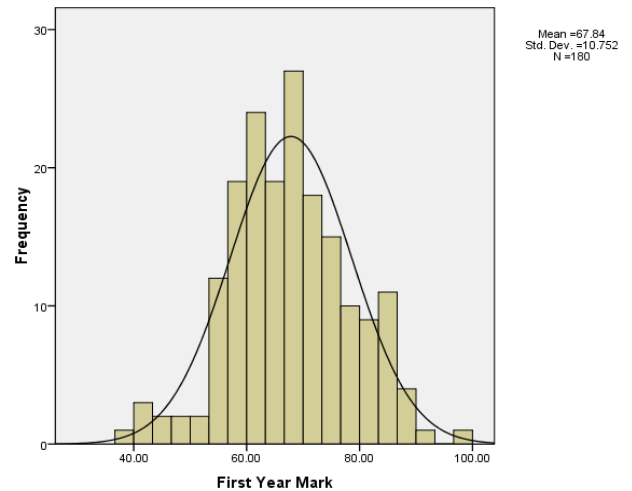


Figure 4.8. Distribution of marks obtained in Integrated Science in first year by students

The mean mark for the distribution is 67.84% and the standard deviation is 10.75. Very few students fell below the 44% pass mark. In fact, 3.3% failed, and about 30.5% fell actually below the average of the distribution.

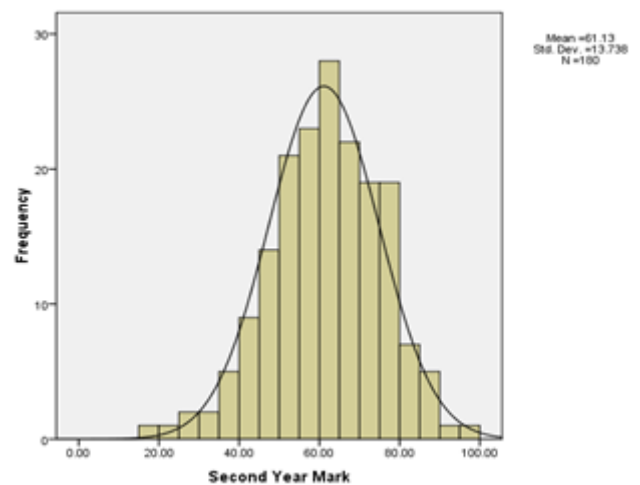


Figure 4.9 Distribution of marks obtained in Integrated Science in second year by students

The average mark is 61.13% and the standard deviation is 13.74. More students fell below the pass mark in comparison with the first year performance showing a decline in second year performance over performance in first year. This has also reflected in the average score dropping from 67 to 61 accounting for about 10.8% decline in performance in terms of the average score.

TABLE 4.7. MODEL SUMMARY FOR INTEGRATED SCIENCE

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.677	.458	.455	10.13861

- a. Predictors: (Constant), first year mark
- b. Dependent Variable: second year mark

The degree of association between the dependent and the independent variable is moderate; the correlation coefficient R is given as 0.677 in table 4.7. The model accounts for 45.8% of total variation in the data. This implies that 45.8% the variance in second year scores is explained by students' first year scores. This percentage is quite significant as it means the academic performance variable has effect than the socio-economic and demographic variables.

TABLE 4.8 COEFFICIENTS FOR INTEGRATED SCIENCE

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
1 (Constant)	2.445	4.840		.505	.614
First year mark	.865	.070	.677	12.275	.000

- a. Dependent Variable: second year mark

The t-value given in table 4.5 is 0.505 and the level of significance is 0.614. This indicates that the effect of students' first year performance on their second year performance is not due to likely random chance, but rather a repeatable cause.

TABLE 4.9. RESIDUAL STATISTICS FOR INTEGRATED SCIENCE

Residuals Statistics					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	35.3194	87.2260	61.1333	9.30187	180
Residual	-3.36216E1	24.48992	.00000	10.11025	180
Std. Predicted Value	-2.775	2.805	.000	1.000	180
Std. Residual	-3.316	2.416	.000	.997	180

Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	35.3194	87.2260	61.1333	9.30187	180
Residual	-3.36216E1	24.48992	.00000	10.11025	180
Std. Predicted Value	-2.775	2.805	.000	1.000	180
Std. Residual	-3.316	2.416	.000	.997	180

- a. Dependent Variable: second year mark

From table 4.9 the minimum predicted value is 35.3% and the maximum predicted value is 87.2%. The grades assigned to these values are F9 and A1 respectively according to the grading scheme of West African Examination Council (WAEC) for West African Senior Secondary Certificate Examination (WASSCE). F9 is the worst grade and per this model some students are likely to have this grade. Some students are also likely to have the best grade in Integrated Science as predicted by this model. From the details of analyzed results in table 4.17 at page 74, there is rather a massive decline in performance from the lower grades to the higher grades. The percentage of students who had grade A1 in first year dropped from 16.7% to 7.8% recording 8.9 % drop. This decline is noted with other grades also bringing the percentage of students that failed from 3.3% to 11.7%.

D. Regression Analysis of Students' Scores in Social Studies

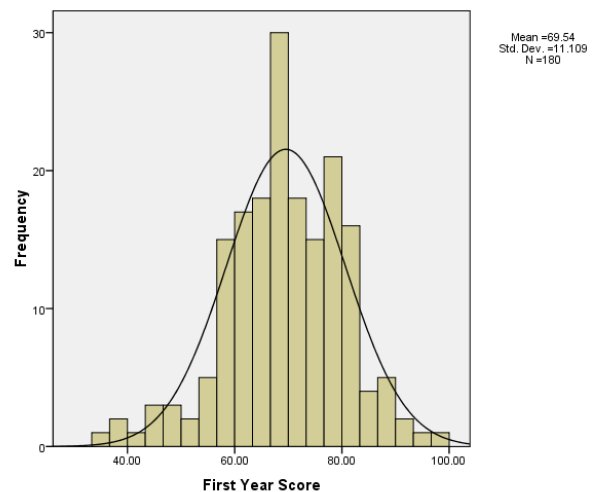


Figure 4.11. Distribution of marks of students in Social Studies in first year

The average score of the distribution is 69.54% and the standard deviation is 11.109. From the chart (fig. 4.11) it is observed that in spite of the high average score some

students' scores fell below the pass mark. Greater number of students scored above 60%.

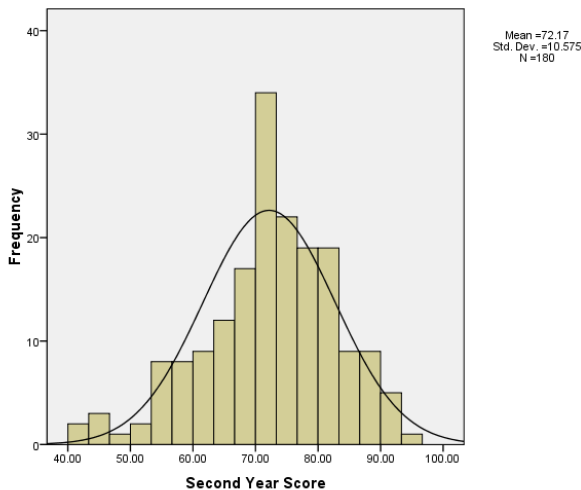


Figure 4.12. Distribution of marks of students in Social Studies in second year

The average score for the distribution is 72.17% and the standard deviation is 10.58. There is a very good improvement in performance as failure rate is reduced significantly. Averagely the distribution recorded about 4.7% increase in performance.

TABLE 4.10. MODEL SUMMARY FOR SOCIAL STUDIES

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.719	.516	.514	7.37468

a. Predictors: (Constant), first year score

b. Dependent Variable: second year score

The degree of association between the dependent and the independent variable is moderate; the correlation coefficient R is given as 0.719 in table 4.10. The model accounts for about 52% of total variation in the data (R Square is 0.516). This means about 52% of the variance in marks of students in Social Studies in second year is explained by the marks students obtained in first year. This percentage is very significant in that it is an indication that students' past academic performance affects their future performance far more than demographic and socio-economic variables.

TABLE 4.11. COEFFICIENTS FOR SOCIAL STUDIES

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	24.604	3.494		7.042	.000
	First year score	.684	.050	.719	13.786	.000

a. Dependent Variable: second year score

The t-value from table 4.11 is 7.042 indicating that the effect of students' first year performance on their second year performance in Social Studies is not due to likely random chance but is due to underlying repeatable cause.

TABLE 4.12. RESIDUAL STATISTICS FOR SOCIAL STUDIES

Residuals Statistics					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	48.5457	90.9572	72.1722	7.59924	180
Residual	-2.36471E1	14.98480	.00000	7.35405	180
Std. Predicted Value	-3.109	2.472	.000	1.000	180
Std. Residual	-3.207	2.032	.000	.997	180

a. Dependent Variable: second year score

Statistics given in table 4.12 show that 48.5% is the minimum predicted value, and about 91% is the maximum predicted value. These values correspond to E8 and A1 respectively in the WAEC grading scheme for WASSCE. It is not an indication that all students will pass in Social Studies because the details of analyzed results in table 4.17 at page 81 shows that 2 students failed. The minimum value also shows that some students have not gone too far away from the failure zone and hence some more work needs to be done in this subject area also. However, a closer look at the results indicates massive improvement in second year performance over the first year performance. There is a massive move from lower grades to higher grades. In fact,

with the failure rate dropping from 3.3% to 1.1%, one can confidently estimate a 0% fail in the final examination.

E. Correlation Analysis of Students' Performance at BECE and SHS

TABLE 4.13. CORRELATION BETWEEN PERFORMANCE IN BECE AND SHS

Correlations		
	BECE	SHS
<i>B E C E</i> Pearson Correlation	1	.518**
<i>C E</i> Sig. (2-tailed)		.000
<i>N</i>	180	180
<i>S H S</i> Pearson Correlation	.518**	1
<i>S</i> Sig. (2-tailed)	.000	
<i>N</i>	180	180

** . Correlation is significant at the 0.01 level (2-tailed).

As indicated in table 4.13 there is a significant degree of correlation between BECE result and SHS performance. The correlation coefficient is 0.518, and it means students' performance in BECE to a large extent determines their performance at SHS. The fact that the correlation coefficient is not exactly 1(perfect correlation), is an indication that there exist other factors that contribute to performance at the SHS level, most of which had been addressed in literature.

V. DISCUSSION

A. Performance of Students in Mathematics

Students' performance in Mathematics in first year was fairly good on the average as the mean mark was 61.46. The standard deviation however gave a clearer view of the performance. A standard deviation of 14.39 shows how the individual student's mark is widely dispersed from the mean, indicating that there were some students who fell far below the average mark (21%).

This performance further declined in the second year bringing the average and the standard deviation to 48.98 and 13.73 respectively. This average clearly fell below 50.0 and this means many more students could not obtain grade D7. 52.8% of the students fell below the average mark that corresponds to D7 as indicated in the following table.

TABLE 5.1 DETAILS OF STUDENTS' PERFORMANCE IN MATHEMATICS.

Year One				Year Two			
Grade	No.	Cum . Freq .	Perce nt	Grade	No.	Cum . Freq .	Perce nt
A1	20		11.1	A1	4		2.2
B2	12	32	6.7	B2	6	10	3.3
B3	15	47	8.3	B3	4	14	2.2
C4	24	71	13.3	C4	9	23	5.0
C5	27	98	15.0	C5	12	35	6.7
C6	25	123	13.9	C6	20	55	11.1
D7	19	142	10.6	D7	30	85	16.7
E8	18	160	10.0	E8	32	117	17.8
F9	20	180	11.1	F9	63	180	35.0
Total Percent			100.0	Total Percent			100.0

The pattern and trend of performance of students in Mathematics extracted from their academic data are indicative of an alarming decline which will eventually reflect in the final examination if nothing is done about it. Better administrative, pedagogical and timely decisions can be taken based on the knowledge discovered for improved performance in Mathematics.

B. Performance of Students in English Language

Performance in English Language in first year was generally good. 92.3% had a score of 50% and above while 7.7% fell below the 50% mark. This performance, however, declined in the second year recording 22.1% falling below the average mark. This trend in performance can be reversed if teachers concentrate efforts on teaching the subject to help improve performance.

TABLE 5.2. DETAILS OF STUDENTS' PERFORMANCE IN ENGLISH LANGUAGE.

Year One				Year Two			
Grade	No.	Cu m. Fre q.	Perce nt	Grade	No.	Cu m. Fre q.	Perce nt
A1	2		1.1	A1	3		1.6
B2	7	9	3.9	B2	14	17	7.8
B3	18	27	10.0	B3	14	31	7.8
C4	24	51	13.3	C4	21	52	11.7
C5	41	92	22.8	C5	33	85	18.3
C6	44	136	24.4	C6	29	114	16.1
D7	30	166	16.8	D7	26	140	14.6
E8	3	169	1.6	E8	16	156	8.8
F9	11	180	6.1	F9	24	180	13.3
Total Percent			100.0	Total Percent			100.0

C. Performance of Students in Integrated Science

Averagely, performance in Integrated Science was good. In the first year, 91% scored 50% and above, leaving 9% which fell below the average mark. Second year saw a decline in performance as the number of students who fell below the average increased from 9% to 19.5%. The number of students who had scores above average dropped from 91% to 80.5% as shown in table 5.3. This fall has also reflected in the class average dropping from 67.84% to 61.13%.

TABLE 5.3 DETAILS OF STUDENTS' PERFORMANCE IN INTEGRATED SCIENCE.

Year One				Year Two			
Grade	No.	Cum. Freq.	Percent	Grade	No.	Cum. Freq.	Percent
A1	30		16.7	A1	14		7.8
B2	18	48	10.0	B2	19	33	10.6
B3	29	77	16.1	B3	18	51	10.0
C4	30	107	16.7	C4	22	73	12.2
C5	33	140	18.3	C5	28	101	15.5
C6	23	163	12.8	C6	23	124	12.7
D7	8	171	4.4	D7	21	145	11.7
E8	3	174	1.7	E8	14	159	7.8
F9	6	180	3.3	F9	21	180	11.7
Total Percent			100.0	Total Percent			100.0

D. Performance of Students in Social Studies

Performance in Social Studies in the first year was averagely good, recording a class average of 69.54%. However, 6.1% fell below 50% leaving 93.9 % scoring above 50%. Remarkably, this performance improved in the second year by recording 96.7% scoring above 50% and the remaining 3.3% below 50%, as indicated in table 5.4. More work need to be done to push the few students who scored below 50% up.

TABLE 5.4 DETAILS OF STUDENTS' PERFORMANCE IN SOCIAL STUDIES.

Year One				Year Two			
Grade	No.	Cum. Freq.	Percent	Grade	No.	Cum. Freq.	Percent
A1	32		17.8	A1	43		23.9
B2	30	62	16.7	B2	33	76	18.3
B3	28	90	15.5	B3	42	118	23.3
C4	34	124	18.8	C4	26	144	14.4
C5	26	150	14.4	C5	12	156	6.7
C6	16	166	8.8	C6	15	171	8.3
D7	3	169	1.6	D7	3	174	1.6
E8	5	174	2.8	E8	4	178	2.2
F9	6	180	3.3	F9	2	180	1.1
Total Percent			100.0	Total Percent			100.0

Inferentially, students 'performance declined in the second year significantly in Mathematics, English, and Integrated Science. This decline cut across all the grades; more students had higher grades in the first year but this could not be maintained or improved in the second year as the number of students who scored higher grades in the first year reduced tremendously. Significant improvement was realized in Social Studies in the second year but there remained a lot more effort to be exerted in order to achieve a 100% average and above average performance.

E. The Degree of association between students' performance in BECE and SHS

There is a significant degree of correlation between BECE result and SHS performance. The correlation coefficient 0.518 shows that students' performance in BECE affects their performance at SHS moderately. What this means is that the poor performance of students in SHS cannot be blamed entirely on the grades they obtained in BECE. The fact that the correlation coefficient is not exactly 1(perfect correlation), is an indication that there exist other factors that contribute to performance at the SHS level, most of which had been addressed in literature.

This finding agrees with literature; (Bharadwaj and Pal., 2011) undertook a study in which Bayesian categorization was used to assess learner accomplishment. 300 learners were selected as sample of which 226 were males and 74 were females. A total number of 17 characteristics were considered to ascertain which of them have impact on the learner accomplishment in Bachelor of Computer Application program in an Indian higher institution of learning called Dr. R. M. L. Awadh University. Questionnaire and institution's database were used to gather statistics on learner's scholastic, personal and social and financial characteristics. Learner's scores were gathered from the institution's assessment unit. Learner's rating in senior secondary exam, residence, channel of receiving tutorial, prerequisite of mother, other addiction of learner, family yearly earning and status of leaner's family were discovered to be extremely linked with scholastic accomplishment of the learner. The study concluded that scholastic accomplishment of learners does not always depend on the learner's own endeavor.

Also, (Kotsiantis et al., 2004) employed several step-by-step methods of DM for predicting the accomplishment of computer science students in distance learning course in a higher institution. For every individual learner, a number of personal characteristics e.g. sex, age, marital status and accomplishment qualities such as scores were used as parameters of a dual classification of pass or fail. The finest result was achieved by a Naive Bayes method which gave precision of 74%. In addition, discovery was made that grades obtained in former institute of scholastic accomplishment has much greater influence than personal characteristics.

Consequently, it can be concluded that performance in the BECE is a determinant of performance in SHS and by implication the CSSPS is significantly a cause of decline in students' performances.

F. Degree of Correlation between Students' First and Second Year Performance

TABLE 5.5 CORRELATION COEFFICIENTS AND T STATISTICS

	<i>Correlation Coefficient (R)</i>	<i>R Square</i>	<i>T Statistic</i>
<i>Mathematics</i>	0.558	0.311	8.962
<i>English Language</i>	0.396	0.157	5.749
<i>Integrated Science</i>	0.677	0.458	12.275
<i>Social Studies</i>	0.719	0.516	13.786

The correlation coefficient in each case in the table above is quite significant. The correlation coefficient measures the degree of association between students' performance in first and second year. The closer these coefficients are to 1, the stronger the bond between the first and second year performance. It is therefore clear from the table that there is a close association between the two performances in all the subjects under study except English Language which shows a weak association. It can therefore be concluded that first year performance has a significant effect on second year performance.

The effect of the first year performance is further explained by the R Square. The R Square explains the variation in second year performance which is caused by first year performance, and in each case the percentage variation is quite significant. As presented in the table, for Mathematics $R^2 = 31\%$, for English Language $R^2 = 16\%$, for Integrated Science $R^2 = 46\%$, and for Social Studies $R^2 = 52\%$.

The association between the first year and second year performances and variation in performance are also validated by the t Statistic. The farther the t statistic is from 0, the more valid the R^2 . The implication is that the variation in performance in first and second year is not due to likely random chance but rather there is underlying repeatable cause.

G. Prediction of Students' Performance in WASSCE

Regression models were built to predict students' performance in WASSCE. The four core subject areas under study each have a predictive model as presented in the methodology.

TABLE 5.6. PREDICTED SCORES

	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Standard Deviation</i>
<i>Mathematics</i>	26.9187	65.7578	48.9778	7.6555
<i>English Language</i>	37.7116	70.7460	58.0389	4.64345
<i>Integrated Science</i>	35.3194	87.2260	61.1333	9.30187
<i>Social Studies</i>	48.5457	90.9572	72.1722	7.59924

1) Prediction of Scores in Mathematics

From table 5.6, the minimum score predicted for Mathematics by the model is about 30%, and it is indicative of students not only falling below average score but failing as well. This situation is very bad in the sense that there is too much gap between this value and the pass mark and with the standard deviation approximated as 7.66; the probability of many students failing is high. This prediction reflects students' second year performance which recorded 35.00% fail, and about 52.80% falling below the average mark of 50%.

The maximum predicted value is 65.76%, which can be graded as C4 by the West African Examination Council (WAEC) scheme for West African Senior Secondary Certificate Examination (WASSCE). The mean is predicted as 48.98%, which comparatively is almost the same as the mean for the distribution of students mark in the second year. It is however lower than the mean for the distribution of students marks in the first year (mean = 61.5%), and it is an indication of a serious fall in performance. It can therefore be concluded that students are not performing high in Mathematics and there is indication of further decline as predicted by the model.

2) Prediction of Scores in English Language

Statistics from table 5.6 show that the predicted minimum, maximum, and mean values are 37.71%, 70.74%, and 58.03% respectively. The standard deviation for the predicted values is 4.64. Comparing the predicted mean with the mean of the distribution of students marks in both first and second year (figure 4.5 and 4.6), it can be seen that they are almost the same (59.69 and 58.04). This statistic reveals the reliability of the regression model. Given the minimum predicted value of 37.71% is indicative of the propensity of some students failing in English Language. In fact, this reflects students' performance in second year which recorded 13.3% failure (F9), about 52.8% falling below the average mark of 58.04%. The maximum predicted value can be converted to a grade B3 according to WAEC grading scheme for WASSCE. This reflects students' second year performance which recorded much decline. This model indicates a better performance in English Language when compared with Mathematics pointing to more work that needs to be done in Mathematics to remedy the situation.

3) *Prediction of Scores in Integrated Science*

From table 5.6, the minimum predicted value is 35.34% and the maximum predicted value is 87.23%. The grades assigned to these values are F9 and A1 respectively according to the grading scheme of West African Examination Council (WAEC) for West African Senior Secondary Certificate Examination (WASSCE). F9 is the worst grade and per this model some students are likely to have this grade. Some students are also likely to have the best grade in Integrated Science as predicted by this model. From the details of analyzed results in table 4.17 at page 74, there is rather a massive decline in performance from the higher grades to the lower grades. The percentage of students who had grade A1 in first year dropped from 16.7% to 7.8% recording 8.9 % drop. This decline is noted with other grades also, bringing the percentage of students that failed from 3.3% to 11.7%. It is therefore certain that even though the model predicted grade A1 as the maximum, only very few students are likely to obtain it. The model predicted 61.13% as the mean score and 9.30 as the standard deviation. The rather high standard deviation shows how widely dispersed the individual scores are from the mean. 40.60% had grades attached to scores above the mean while 41.90% had grades below the grades that correspond to scores below the mean in the second year.

4) *Prediction of Scores in Social Studies*

Statistics from table 5.6 show that 48.55% is the minimum predicted value, and approximately 90.96% is the maximum predicted value. These values correspond to E8 and A1 respectively in the WAEC grading scheme for WASSCE. It is an indication that all students will pass in Social Studies. This, however, does not call for complacency because scores of some students definitely will fall below the 50% average. The minimum value also shows that some students have not gone too far away from the failure zone and hence some more work needs to be done in this subject area also. However, a closer look at the results indicates massive improvement in second year performance over the first year performance. There is a significant move from lower grades to higher grades. In fact, with the failure rate dropping from 3.3% to 1.1%, one can confidently estimate a 0% failure rate in the final examination even as the model projected. The mean value predicted is 72.17% and this is consistent with the mean of students' scores in second year which proves the reliability of the model.

VI. SUMMARY, CONCLUSION, AND RECOMMENDATION

A. *Summary*

Mawuli School is one of the renowned second cycle institutions in Ghana, and most especially in the Volta Region of Ghana. In fact, it is referred to as "Volta University" due to its unparalleled performance in the past. In recent times, however, performance does not seem to be what it used to be, even though better than those of some of its counterparts. Comparative analysis of the West African Senior School Certificate Examination of 2013 revealed that Mawuli School placed fourth position in the Volta Region. Out of 1089 candidates who took the examination, 599

students passed in all eight subjects, representing 55% and only 471 qualified for tertiary level, representing 43.5%. This decline is so sharp when compared to 2011 and 2012 results that recorded 90.6% and 89.3% of candidates that qualified for tertiary level.

Stakeholders always expect high performance even though the stock of candidates presented for WASSCE may not be the best, and most often results of poor performance hit them by surprise. Perception among stakeholders of education in Mawuli School is that the stock of students being admitted through the Computerized School Placement System to a very large extent is affecting performance negatively.

With the situation at hand it is worth analyzing students' performance, come out with a model that predicts students' final grade to give all stakeholders a general overview of the final result, so that attention could be given to areas of possible flops ahead of final examination.

Relevant literature was reviewed covering areas such as similar studies conducted by various researchers, the concept of educational data mining, and data mining techniques. The researcher also sought to investigate the effect of the Computerized School Placement System on Students' performance and find the extent of it.

The target population was the second year students of 2012 to 2015 academic year. Data for the study was collected from a secondary source; from students' academic records in Mawuli School, and a sample size of 180 was considered for the study. Three sets of data were collected namely: marks of students in all core subjects in first year, marks of students in all core subjects in second year, and grades of students in all core subjects in BECE.

The data was analyzed using SPSS v16. Simple linear regression analysis was performed to analyze performance and construct prediction models to estimate students' final scores in all core subjects. Also, to investigate the relationship and determine the degree of association between students' performance in SHS and their grades at BECE, a Pearson's correlation analysis was used.

Results of the study were presented in tables, charts, and formulas. It was revealed that there was a decline in students' performance in Mathematics, English Language, and Integrated Science, and a good number of students fell below the pass mark, although some few students were performing high, while Social Studies saw a good improvement in students' performance over performance in the first year by recording a 98.9% pass.

It was also found that there is a close association between students' first year and second year performance, and that the effect of students' first year performance on their second year performance was not due to likely random chance but that the effect had underlying repeatable cause, and hence models were constructed to predict students' final WASSCE scores/grades. For Mathematics, the minimum

mark predicted is 26.9%, which corresponds to grade F9, and the maximum mark predicted is 65.6%, corresponding to grade C4. The average is 48.0%, and this corresponds to grade E8. The minimum mark predicted for English Language is 37.7, which is grade F9; the maximum mark predicted is 70.7%, which is grade B3, and the average is 58.0, which is grade C6. For Science, 35.3%, 87.2%, and 61.1%, corresponding to grades F9, A1, C5 were predicted for the minimum, maximum, and the average respectively. Predictions for Social Studies were 48.5%, 90.6%, and 72.2% as minimum, maximum, and the average mark respectively. These marks represent E8, A1, and B3 respectively. Correlation does exist between the performance in BECE and performance at SHS. The correlation coefficient is 0.518, and it means students' performance in BECE to some extent accounts for their performance at SHS. The implication therefore is that the Computerized School Placement System accounts for the poor performance of students as perceived by stakeholders. The fact, however, is that because the correlation coefficient is not exactly 1(perfect correlation), poor performance cannot be blamed entirely on the Computerized School Placement System, implying there are other factors that contribute to students' poor performance at the SHS level.

B. Conclusion

The research questions stated for the study were fully examined in relation to the results of the study. It was found that few students performed high in Social Studies, Integrated Science, and English Language than in Mathematics even though performances in all core subjects except Social Studies kept declining. Also, it was established from the results that there is a significant degree of association between performance in BECE and SHS. Performances at the two levels were moderately correlated, meaning indeed the grades of students in BECE affect their performance at SHS. This finding agrees with literature; that the factors like students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance (Bharadwaj and Pal, 2011). Also, it was found that past school grades have a much higher impact than demographic characteristics (Kotsiantis et al., 2004).

However, one cannot blame poor performance on the Computerized School Placement System entirely since the correlation coefficient does not show perfect correlation.

Finally, it was clearly seen that there exists significant degree of association between students' first year scores and second year scores. This implies that students' performance in the second year of their program was affected by their performance in the first year, and clearly this effect was not due to likely random chance but had underlying repeatable cause. This also implies that students' terminal performance affects their final results to a large extent. Based on this fact prediction models were constructed to project students' final scores. However, coefficients of determination in all cases indicate that there are other factors that explain the variance

in scores. This is in line with works done by earlier researchers. (Aysha et al. 2010) employed a method of data mining called step-by-step k-means clustering to forecast behaviors of learners in a learners' record comprising class tests, middle of term and end of term exams and projects. This connected statistics would be transferred to the class instructor for the organization of final exam. The investigation helped the instructors to bring down the declining percentage by applying the right actions at appropriate time to enhance learner accomplishment.

Also, (Kotsiantis et al., 2004) applied several DM algorithms to predict the performance of computer science students, and concluded that past school grades have a much higher impact than demographic variables.

C. Recommendations

Based on the findings of the study, it is recommended that:

1. Management and teachers in Mawuli School must find ways to intensify efforts to improve performance in the three core subject areas namely Mathematics, Integrated Science, and English Language.
2. Stakeholders should not have baseless expectations of students' performance but rather consider students' terminal performance as a guide.
3. Management, teachers, and other stakeholders of Mawuli School must find ways to deal with other factors that affect students' performance such as demographic and socio- economic variables like gender, grade level, school location, school type, student type, student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education as revealed in literature.
4. Management must initiate remedial efforts to help weak students immediately they are admitted for their programs in order to help improve their performance far ahead of the WASSCE examination.

VII. ACKNOWLEDGMENT

I would first like to thank the almighty God of heaven for His love, goodness, support, and wisdom He provided me during the entire period of my studies in the university. In fact, I owe the very foundation of my life and success to Him.

My gratitude and respect go to Mr. Dominic Asamoah and Dr. Ussif Najim who through patience and professional expertise supervised this work. I do appreciate you for your careful attention to each of the chapters and all the suggestions and directions you offered me throughout this work.

Special thanks to Rev. S.S. Asieni, the Assistant Headmaster Academic, Mawuli School, for his immense support toward the success of this work.

Finally, kudos to all friends and course mates who shared their rich experiences, ideas, and suggestions to bring

this work to accomplishment. All your contributions are very much appreciated.

REFERENCES

- [1] Adebajo, A.T., Oladokun, V.O., and Charles-owaba, O.E. (2008). "Predicting Student's Academic Performance Using Artificial Neural Network: A Case Study of an Engineering Course". *Pacific Journal of Science and Technology*, 9 (1), 72-79.
- [2] Al-Radaideh, Q., Al-Shawakfa, E., and Al-Najjar, M., (2006). "Mining Student Data Using Decision Trees". The 2006 International Arab Conference on Information Technology (ACIT' 2006).
- [3] Anozie, N., and Junker, B.W. (2006). "Predicting End-of-year Accountability Assessment Scores from Monthly Student Records in an Online Tutoring System". *Educational Data Mining AAAI Workshop* (pp. 1-6). California.
- [4] Arnold, A., Scheines, R., Beck, J.E., and Jerome, B. (2005). "Time and Attention": Students, Sessions, and Tasks. *AAAI2005 Workshop on Educational Data Mining* (pp. 62-66). Pittsburgh.
- [5] Ayesha, S., Mustafa, T., Sattar, A.R., and Khan, M.I. (2010). "Data Mining Model for Higher Education System. *European Journal of Scientific Research*", 43(1), 24-29.
- [6] Ayers E., and Junker B.W. (2006). "Do Skills Combine Additively to Predict Task Difficulty in Eighth Grade Mathematics?" *AAAI Workshop on Educational Data Mining*, pp. 14-20. Menlo Park.
- [7] Baker, R. (2010). "Data Mining for Education". In: McGraw, B., Peterson, P., Baker, E. (eds.) *International Encyclopedia of Education*, 3rd ed. Elsevier, Oxford.
- [8] Baker, R. and Carvalho, A. (2008). "Labeling Student Behavior Faster and More Precisely with TextReplays". *1st International Conference on Educational Data Mining*, pp 38-47.
- [9] Baker, R., Barnes, T., and Beck, J.E. (2008). "Educational Data Mining 2008": *1st International Conference on Educational Data Mining, Proceedings*. Montreal, Quebec, Canada.
- [10] Baker, R., and Yacef, K. (2009). "The State of Educational Data Mining in 2009: A Review and Future Visions". *Journal of Educational Data Mining*, 1 (1), 3-17.
- [11] Beck, J.E., and Woolf, B.P. (2000). "High-level Student Modeling with Machine Learning". *Fifth International Conference on Intelligent Tutoring Systems*, (pp. 584-593). Alagoas, Brazil.
- [12] Berson, A., Smith, S., and Thearling, K. (2011). "An Overview of Data Mining Techniques". Retrieved from <http://www.thearling.com/text/dmtechniques/dmtechniques.htm>
- [13] Bharadwaj, B.K., and Pal, S. (2011). "Data Mining: A Prediction for Performance Improvement Using Classification". *International Journal of Computer Science and Information Security (IJCSIS)*, 9 (4), 136-140.
- [14] Bradley, A.P. (2007). "The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms". *Pattern Recognition*, 30, 1145-1159.
- [15] Calders, T., and Pechenizkiy, M. (2012), "Introduction to the Special Section on Educational Data Mining". *SIGKDD Explorations*, 13 (2) pp.3-6.
- [16] Campbell, J.P., and Oblinger, D.G. (2007). "Academic Analytics". *EDUCAUSE*, Washington, D.C.
- [17] Castro, F., Vellido, A., Nebot, A., and Mugica, F. (2007). "Applying Data Mining Techniques to e-Learning Problems". In Jain, L.C., Tedman, R. and Tedman, D. (eds.), *Evolution of Teaching and Learning Paradigms in Intelligent Environment: Studies in Computational Intelligence*, 62, 183-221. Springer-Verlag.
- [18] Cetintas, A., Si, L., Xin, Y.P., and Hord, C. (2009). "Predicting Correctness of Problem Solving from Low-level Log Data in Intelligent Tutoring Systems". *International Conference on Educational Data Mining*, (pp 230-238). Cordoba, Spain.
- [19] Chakrabarti, S., Thomas P., Nadeau, et al. (2009). "Data Mining". Morgan Kaufmann Publishers, Burlington.
- [20] Chadha A. and Kumar V (2011). "An Empirical Study of the Applications of Data Mining Techniques in Higher Education". *International Journal of Advanced Computer Science and Applications*, 2(3), pp. 80-84.
- [21] Chan, C.C. (2007). "A Framework for Assessing Usage of Web-Based e-Learning Systems". *International Conference on innovative Computing, Information and Control*, Washington, DC, 147- 151.
- [22] Chen, C., Chen, M., and Li, Y. (2007). "Mining Key Formative Assessment Rules Based on Learner Profiles for Web-based Learning Systems". *IEEE International Conference on Advanced Learning Technologies*, Japan, pp 1-5.
- [23] Cooper, M., Stevens, R., Giordani, A., Soller, A and Gerosa, L., (2005). "Developing a Framework for Integrating Prior Problem Solving and Knowledge Sharing Histories of a Group to Predict Future Group Performance". *International Conference on Collaborative Computing: Networking, Applications and Worksharing*, (pp. 1-9). Boston.
- [24] Corbett, A. T. (2001). "Cognitive Computer Tutors: Solving the Two-Sigma Problem". *8th International Conference on User Modeling*.
- [25] Corbett, A.T., and Anderson, J.R., (1995). "Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge, User Modeling and User-Adapted Interaction", 4, 253-278.

- [26] Creswell, J.W (2003). "Research design: qualitative & quantitative approaches". Thousand Oaks, CA: Sage
- [27] Dawson, C. (2002). "Practical Research Methods: a User-friendly Guide to Mastering Research Techniques and Projects". How to Books Ltd. Oxford OX4 IRE, UK.
- [28] Delgado, M., Gibaja, E., Pegalajar, M.C., and Pérez, O. (2006). "Predicting Students' Marks from Moodle Logs using Neural Network Models". International Conference on Current Developments in Technology-Assisted Education, Sevilla, Spain, 586-590.
- [29] Desmarais, M.C., Gagnon, M., and Meshkinfram, P. (2006). "Bayesian Student Models Based on Item to Item Knowledge Structures". Conference on Technology Enhanced Learning, (pp 1-10). Crete, Greece.
- [30] Draper, N.R., and Smith, H., (1998)."Applied Regression Analysis".Wiley.
- [31] Dunham, M. (2003). Data Mining: "Introductory and Advanced Topics". Pearson Education, Upper Saddle River, NJ.
- [32] El-Halees, A. (2008). "Mining Students Data to Analyze Learning Behavior: A Case Study". The 2008 international Arab Conference of Information Technology (ACIT2008) – Conference Proceedings. University of Sfax, Tunisia, Dec 15- 18.
- [33] Espejo, P., Ventura, S., and Herrera, F. (2010)."A Survey on the Application of Genetic Programming to Classification". IEEE Transactions on Systems, Man, and Cybernetics-Part C, 40, (2), 121-144.
- [34] Etchells, T.A., Nebot, A., Vellido, A., Lisboa, P.J.G., and Mugica, F. (2006). "Learning what is Important: Feature Selection and Rule Extraction in a Virtual Course". European Symposium on Artificial Neural Networks, Bruseles, Belgium pp 401-406.
- [35] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). "From Data Mining to Knowledge Discovery: An Overview". Advances in Knowledge Discovery and Data Mining, MIT Press, pp 1-3.
- [36] Feng, M., Beck, J., and Heffernan, N. (2009)."Using Learning Decomposition and Bootstrapping with Randomization to Compare the Impact of Different Educational Interventions on Learning".2nd International Conference on Educational Data Mining, (pp 51-60).Cordoba, Spain.
- [37] Feng, M., Heffernan, N., and Koedinger, K. (2005). "Looking for Sources of Error in Predicting Student's Knowledge". AAAI'05 workshop on Educational Data Mining, (pp 1-8). Amsterdam.
- [38] Freedman, D., Purves, R., and Pisani, R. (2007). "Statistics". 4th Edition. W.W. Norton & Co.
- [39] Freyberger, J., Heffernan, N.T., and Ruiz, C. (2004). Using Association Rules to Guide a Search for Best Fitting Transfer Models of Student Learning". Workshop Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, (pp.1-10). Alagoas, Brazil.
- [40] Garcia, P., Amandi, A., Schiaffino, S., and Campo, M. (2008). "Evaluating Bayesian Networks' Precision for Detecting Student's Learning Styles". Computer and Education Journal, 49, 794-808.
- [41] Gedeon, T.D., and Turner, H.S. (1993). "Explaining Student Grades Predicted by a Neural Network". International conference on Neural Networks, (pp. 609-612). Nagoya.
- [42] Girones, M., and Fernandez, T.A. (2006). "Ariadne, a Guiding Thread in the Learning Process's Labyrinth". International Conference on Current Developments in Technology-Assisted Education, (pp. 287-290).Sevilla.
- [43] Golding, P., and Donalson, O. (2006). "Predicting Academic Performance". Frontiers in Education Conference, (pp 21-26). San Diego, California.
- [44] Guan, J., Nunez, W., and Welsh, J. (2002). "Institutional Strategy and Information Support": The Role of Data Warehousing in Higher Education. Campus-Wide Information Systems, 19 (5), 168-174
- [45] Haddawy, P., Thi, N., and Hien, T.N. (2007). "A Decision Support System for Evaluating International Student Applications". Frontiers in Education Conference, (pp. 1-4). Milwaukee.
- [46] Hamalainen, W., and Vinni, M. (2006). "Comparison of Machine Learning Methods for Intelligent Tutoring Systems". International conference in intelligent tutoring systems, (pp. 525-534). Taiwan.
- [47] Han, J. and Kamber, M. (2001). "Data Mining: Concepts and Techniques". Simon Fraser University, Morgan Kaufmann Publishers, ISBN 1-55860-489-8.
- [48] Hershkovitz, A., and Nachmias, R. (2008). "Developing a Log-based Motivation Measuring Tool". 1st International Conference on Educational Data Mining, (pp. 226-233). Montreal.
- [49] Hijazi, S. T., and Naqvi, R. S. M. M. (2006). "Factors Affecting Student's Performance: A Case of Private Colleges". Bangladesh e-Journal of Sociology, 3, (1). (pp. 97-106).
- [50] IBM. (2012). "What is big data?" Retrieved from 01.ibm.com/software/data/bigdata/
- [51] Ibrahim, Z., and Rusli, D. (2007). "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression". Annual SAS Malaysia Forum, (pp. 1-6). Kuala Lumpur.
- [52] Industrial Research Institute (2010). "Research Management". Industrial Research Institute, Michigan.
- [53] Johnson, B., and Christensen, L., (2010). "Educational Research: Quantitative, Qualitative, and Mixed Approaches". UK: SAGE.
- [54] Kay, J., Maisonneuve, N., Yacef, K., and Zaiane, O.R. (2006). "Mining Patterns of Events in Students' Teamwork Data". Proceedings of Educational Data Mining Workshop, (pp 1-8). Taiwan.

- [55] Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., and Haydock, M. (2012). Analytics: "The Widening Divide". MIT Sloan Management Review, 53(2), 1-22
- [56] Kotsiantis, S.B., and Pintelas, P.E., (2005). "Predicting Students' Marks in Hellenic Open University". IEEE international Conference on Advanced Learning Technologies, (pp. 664-668). Washington, DC.
- [57] Kotsiantis, S., Pierrakeas, C., and Pintelas, P., (2004). "Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques". Applied Artificial Intelligence, Vol. 18, No. 5, pp. 411-426.
- [58] Larose, D., (2006). "Data Mining: Methods and Models". John Wiley.
- [59] Malik, K., and Ranjan, J., (2007). "Effective Educational Process: A Data Mining Approach". VINE: The Journal of Information and Knowledge Management Systems, 37(4), 502-515.
- [60] Martinez, D., (2001). "Predicting Student Outcomes Using Discriminant Function Analysis". Meeting of the Research and Planning Group, pp. 1-22. Lake Arrowhead, CA.
- [61] McDonald, B., (2004). "Predicting students' success". Journal for Mathematics Teaching and Learning, (pp. 1-14).
- [62] Merceron, A., and Yacef, K. (2008). "Interestingness Measures for Association Rules in Educational Data". International Conference on Educational Data Mining, (pp. 57-66) Montreal, Canada.
- [63] Minaei-bidgoli, B., Kashy, D.A., Kortmeyer, G., and Punch, W.F. (2003). "Predicting Students' Performance: An Application of Data Mining Methods with an Educational Web-based system". International Conference on Frontiers in Education, (pp.13-18). Colorado, USA.
- [64] Moriana, J.A., Alos, F., Alcalá, R. Pino, M. J., Herruzo, J., et al. (2006). "Extra-Curricular Activities and Academic Performance in Secondary Students". Electronic Journal of Research in Educational Psychology, 4 (1), 35-46.
- [65] Myller, N., Suhonen, J., and Sutinen, E. (2002). "Using Data Mining for Improving Web-Based Course Design". International Conference on Computers in Education, (pp. 959- 964). Washington.
- [66] Nebot, A., Castro, F., Vellido, A., and Mugica, F. (2006). "Identification of Fuzzy Models to Predict Students' Performance in an e-Learning Environment". International Conference on Web-based Education, (pp. 74-79). Puerto Vallarta.
- [67] Nemati, H., and Barko, C. (2004). "Organizational Data Mining (ODM): An Introduction". In H. Nemati and C. Barko (Eds.), Organizational Data Mining (pp. 1-8). London: Idea Group Publishing.
- [68] Ogor, E.N. (2007). "Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques". Electronics, Robotics and Automotive Mechanics Conference, (pp. 354-359). Washington, DC.
- [69] Pardos Z.; Heffernan N.; Anderson B.; and Heffernan C., (2006). "Using Fine-Grained Skill Models to Fit Student Performance with Bayesian Networks". Proceedings of 8th International Conference on Intelligent Tutoring Systems. Taiwan.
- [70] Pardos, Z., Heffernan, N., Anderson, B., and Heffernan, C. (2007). "The Effect of Model Granularity on Student Performance Prediction Using Bayesian Networks". International Conference on User Modeling, (pp. 435-439). Corfu, Greece.
- [71] Pardos, Z., Beck, J.E., Ruiz, C., and Heffernan, N. (2008). "The Composition Effect: Conjunctive or Compensatory? An Analysis of Multi-Skill Math Questions in ITS". International Conference on Educational Data Mining, (pp. 147-156). Montreal.
- [72] Pritchard, D., and Warnakulasooriya, R. (2005). "Data from a Web-based Homework Tutor Can Predict Student's Final Exam Score". World Conference on Educational Multimedia, Hypermedia and Telecommunications, (pp. 2523-2529). Chesapeake.
- [73] Rajagopalan, S. P., and Shyamala, K., (2006). "Data Mining Model for a Better Higher Educational System". Information Technology Journal, 5 (3), 560-564.
- [74] Romero, C. and Ventura, S. (2007). "Educational Data Mining: A survey from 1995 to 2005". Expert Systems with Applications, 33, 135-146.
- [75] Romero, C., Ventura, S., Hervás, C., and Gonzales, P. (2008). "Data Mining Algorithms to Classify Students". International Conference on Educational Data Mining, (pp. 8-17). Montreal, Canada.
- [76] Sargenti, P., Lightfoot, W. and Kehal, M. (2006). "Diffusion of Knowledge in and through Higher Education Organizations". Issues in Information Systems, 3 (2), 3-8.
- [77] Saunders, M., (2003). "Research Methods for Business Students". Pearson Education, South Africa.
- [78] Scheines, R., Sprites, P., Glymour, C., and Meek, C. (1994). "Tetrad II: Tools for Discovery". Lawrence Erlbaum Associates: Hillsdale, NJ.
- [79] Shangping, D., and Ping, Z. (2008). "A Data Mining Algorithm in Distance Learning". International Conference on Computer Supported Cooperative Work in Design, (pp. 1014-1017), Xian.
- [80] Strube, M.J. (1985). "Combining and Comparing Significance Levels from Non Independent Hypothesis Tests". Psychological Bulletin, 97, 334-341.
- [81] Tair, M.M.A., and El-Halees, A.M., (2012). "Mining Educational Data to Improve Students' Performance". International Journal of Information Technology Research, 2(2).
- [82] Thomas, E.H., and Galambos, N. (2004). "What Satisfies Students? Mining Student-Opinion Data with Regression and Decision Tree Analysis". Research in Higher Education Journal, 45 (3), 251-269.

- [83] Walonoski, J., and Heffernan, N.T. (2006). "Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems". Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 382-391.
- [84] Walters, Y.B., and Soyibo, K. (2001). "An Analysis of High School Students' Performance on Five Integrated Science Process Skills". Research in Science & Technical Education, 19 (2), 33 – 145.
- [85] Wang, A.Y., and Newlin, M.H. (2002). Predictors of Web-Based Performance: The Role of Self-Efficacy and Reasons for Taking an On-Line Class". Computers in Human Behavior Journal, 18, 151-163.
- [86] Want, T., and Mitrovic, A. (2002). "Using Neural Networks to Predict Student's Performance". International Conference on Computers in Education, (pp. 1-5). Washington, DC.
- [87] Witten, I.H. and Frank, E. (1999). "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations". Morgan Kauffman, San Francisco, CA.
- [88] Wu, A., and Leung, C. (2002). "Evaluating learning behavior of Web-Based Training (WBT) using Web log". International Conference on Computers in Education, (pp. 736-737). New Zealand.
- [89] Yu, C.H., Jannasch-pennell, A., Digangi, S., and Wasson, B. (1999). "Using On-line Interactive Statistics for Evaluating Web-based Instruction". Journal of Educational Media International, 35, 157-161.
- [90] Zinn, C., and Scheuer, O. (2006). Getting to Know Your Students in Distance Learning Contexts". 1st European Conference on Technology Enhanced Learning, (pp. 437-451).