A REGRESSION MODEL TO PREDICT THE RISK OF INCOMPLETE GRADING OF STUDENT ASSESSMENTS IN HIGHER EDUCATION INSTITUTIONS

By

GEOFFREY OCHIENG OKOTH

KCA/07/04424

MASTER OF SCIENCE IN DATA ANALYTICS

KCA UNIVERSITY

2023

A REGRESSION MODEL TO PREDICT THE RISK OF INCOMPLETE GRADING OF STUDENT ASSESSMENTS IN HIGHER EDUCATION INSTITUTIONS

By

GEOFFREY OCHIENG OKOTH

KCA/07/04424

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A DEGREE OF MASTER OF SCIENCE IN DATA ANALYTICS IN THE SCHOOL OF TECHNOLOGY AT KCA UNIVERSITY

OCTOBER 2023

DECLARATION

I declare that this thesis is my original work and has not been previously published or submitted elsewhere for award of a degree. I also declare that this contains no material written or published by other people except where due reference is made, and author duly acknowledged.

This research paper is wholly original with no submissions for awards to any other universities.

Student Name: Geoffrey Ochieng Okoth **Registration Number:** KCA/07/04424

Signature................................... Date...

This research paper proposal has been submitted for consideration with my approval as KCA University Supervisor.

Dr. Lucy Mburu Waruguru **Date**:

ABSTRACT

The accuracy and completeness of student assessment data are paramount in higher education institutions, serving as a cornerstone for informed decision-making, equitable education, and student success. However, the issue of incomplete grading, where grades for assessments are missing or inaccurate, poses a significant challenge. This research presents a regression model designed to predict the risk of incomplete grading of student assessments in higher education institutions. By leveraging historical data, the model identifies factors contributing to incomplete grading, such as grading errors, data entry issues, and technological challenges. Moreover, it examines the consequences of incomplete grading, encompassing student well-being, academic performance, and institutional accountability. The model, built using a comprehensive dataset and machine learning techniques, serves as a valuable tool for educational institutions to proactively address and mitigate the issue of incomplete grading. The research targeted a population of 367 and higher education students from Kenyan universities. Online questionnaires were used to get data from the respondents and SPSS was used to convert data into numerical values. The data collected was analyzed using python data analysis tool to identify patterns and generate the model. The source-code was written in Python. The ANOVA statistic showed that the independent variables are significant to the dependent variable. Subsequently, the independent variables in the study have a significant impact on the dependent variable of Sustainable prediction of incomplete grading. The findings of the research are significant to the education sector as it adds knowledge that will help guide the institutions on how to manage missing marks.

***Keywords*:** Missing marks, multiple linear regression, python, assessment, eLearning.

ACKNOWLEDGEMENT

First and foremost, I would like to thank God Almighty for giving me the strength, good health, will, and for sustaining me throughout my course of study. I extend my appreciation and gratitude to all those that contributed immensely towards completion of this research project.

I especially would like to thank my family, and especially my mother for her belief in me and her tireless encouragement during my study. This work is dedicated to them for the sacrifice, care, love, concern, encouragement, support and enthusiasm. I owe a debt of gratitude to my friends and work colleagues who sacrificed time and gave me invaluable support that saw me through the challenging period. A special thanks to my loving family who were especially supportive in listening to my thoughts and helped me work out logistical details throughout this process.

I want to thank my supervisor Dr. Lucy Mburu, who guided, encouraged, and led me through the journey to completing this proposal.

I want to express my gratitude to everyone who, in some way or another, helped me to attempt and complete this research proposal.

ACRONYMS AND ABBREVIATIONS

MATLAB	MATrix LABoratory
MPGRNN	Multilayer Perceptron and Generalized Regression Neural Network
UTAUT	Unified Theory of Acceptance and Use Technology
UNWE	University of National and World Economy
RF	Random Forest
DT	Decision Tree
SVM	Support Vector Machine
ML	Machine Learning
BNN	Bayesian Belief Network
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
MAE	Mean Absolute Error

GLOSSARY

TERM	DEFINITION
Adoption and integration	Technology that has been accepted and put to use by an organization.
Model	To create a representation of something on a small scale so as to base yourpredictions of the future outcome.
Mobile Technology	Technology that goes where the user goes.
Machine learning	the use and development of computer systems that are able to learn andadapt without following explicit instructions, by using algorithms andstatistical models to analyze and draw inferences from patterns in data.
Endpoint Security	Endpoint Security refers to protecting various end-user devices like laptops,smartphones, or tablets. Those endpoints serve as points of access to the corporate network and sensitive data.
Regressions analysis	Regression analysis is a set of statistical methods used for the estimation ofrelationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.
Predictive modelling	Predictive modeling is the general concept of building a model that iscapable of making predictions. Typically, such a model includes a machine learning algorithm that learns certain properties from a training dataset in order to make those predictions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

In universities, grading is an essential component of the learning process since it enables learners to measure their understanding and competence in a given subject (Ruiz-Jimenez et al., 2022). However, missing marks can lead to incomplete grading, which may lead to confusion and dissatisfaction for both students and instructors. Based on the audit report produced by the Commission for University Education (CUE), it's revealed that some universities had a persisting problem of missing student marks which led to delayed graduation for students in all levels, too many procedures used for obtaining their examination results and transcripts and validity of degrees awarded (CUE, 2019). Reports on errors in recording missing marks and loss of students' examination data were also recorded which would lead to student's frustration when they miss on graduation (Ondari, 2019).

Missing marks, in essence, refer to the troubling instances where recorded grades for student assessments are either incomplete, inaccurate, or entirely absent from educational records. It is a multifaceted challenge that not only poses a conundrum for educational institutions but raises significant questions about the fairness, equity, and integrity of educational assessment. As a society that values education as a cornerstone of progress, understanding and addressing the issue of missing marks is of paramount importance. This study will embark on an exploration of missing marks, delving deep into its intricate web of causes, consequences, and potential solutions. It seeks to unravel the multifaceted nature of this issue, shed light on the myriad factors contributing to its occurrence, and offer insights into the far-reaching consequences it imparts on students, educators, and the educational system as a whole.

Students from lower income brackets have a higher risk of not completing school due to incomplete grading. School misbehavior, school mobility, grade retention, homework completion, school safety, attendance, academic self-efficacy, academic engagement, and academic expectations were all uniquely related to grade point average and experiencing two of the risk factors was predictive of academic failure. The developed sessions were based on self-determination theory and motivational interviewing skills. The sessions include strategies to help with organization, time-management, and self-advocacy. These sessions are also designed to create a support system with peers and a trusted adult. Over the course of six small groups students will be exposed to multiple skills. Once they learn about these different skills, students will be guided to creating a plan and using skills that help them achieve academic success throughout their school career (Mroczek et al., 2022).

Higher education is essential for learners to equip themselves with technical skills through studying so as to contribute to the society. Most of these students graduate each year, but many of them fail the course and have to retake it to advance to the next level or complete their degree. According to the evidence, a large number of these students struggle with depression and poor self-esteem (Feng et al., 2022). This significantly increases faculty workload and increases costs for universities (Keane et al., 2022). Academic advancement of students and educational institutions may be significantly impacted by incomplete grading on student assessments. Predictive modeling approaches have been employed to predict the chances of incomplete grading based on numerous elements and patterns seen in past data in order to address this issue pro-actively. This background investigation intends to explore the current literature and study on models of prediction for forecasting incomplete grading on student evaluations.

From a global standpoint, if a sizable proportion of students drop out of college owing to academic failure, not only will universities' reputations be damaged, but society goals will also be compromised (Tamada et al., 2022). In order to reduce the possibility of wasting public funds on students who ultimately fail, it is urgently necessary to create an accurate method of forecasting academically at-risk university students (Jongile, 2022).

Nearly 30% of college students in the United States have taken at least one or more courses, either virtually or physically, and this trend is expected to increase in the future. Of particular note is the assessment of student learning success. This includes how teachers design and implement formative and summative assessments to measure student progress and learning, provide effective feedback, and ensure academic integrity (Goss, 2022).

Active student participation is encouraged through effective learning assessment. Data were collected from a sample of Alberta post-secondary teachers using an online survey. To assess learning potential, the questionnaire used three criteria, Learning-oriented assessment, assignments as learning assignments (genuine), self-assessment and peer assessment, and feedback. As a result, only 3 of 15 full-fledged tasks were practiced by more than 30% of educators. The findings also showed that teachers' opinions on the use of student interaction and feedback varied. These results indicate that learners' involvement in assessment strategies is limited and can affect learning. Teachers are encouraged to use professional development to understand different methods of authentic assessment and how to optimize active student participation in feedback (Rahman et al., 2022).

Incomplete data has been modelled using mathematical algorithms that have been able to compensate for missing marks (Sandip, 2021). According to (Muchai et al., 2021) cases of missing marks have been resolved using mathematical algorithms where data collected was through mixed methods e.g. qualitative and quantitative and the findings of the two methods were combined and results were found to have a positive delivery on post examination service delivery.

Clustering analysis is a main technique used to solve the problems that exist in data mining. It aims to classify data points into several groups called clusters so that similar elements are clustered into one group, while different elements are separated from each other. Because clustering can be used to deeply mine internal and possible knowledge, rules, and patterns, it has been applied to many practical fields, including data mining, pattern recognition, machine learning, information retrieval, and image analysis. Clustering has been widely and intensely studied by the data mining community in recent years. Many clustering algorithms have been proposed: K-means, fuzzy C-means, affinity propagation, Gaussian mixtures, etc. These algorithms perform well on many occasions, but they cannot deal with datasets with arbitrary shapes, and the determination of hyper parameters and clustering centers is also difficult for clustering algorithms (Gao et al., 2022).

Knowing that a single grade should not define a student’s success, higher education institutions created various grading policies that focused on student success (Kleinman et al., 2018). For example, to address students’ trepidation with grades and to support their academic success, institutions have implemented plus/minus grading structures, grade extension options, course withdrawals, and multiple grade forgiveness attempts for a course 2 (Kleinman et al., 2018).

Kanetaki et al. (2022) state that grading is a critical component of the academic experience in universities, providing students with a measure of their knowledge and skills in a particular subject. However, missing marks can occur, which can lead to confusion and frustration for both students and instructors and these results to high student attrition rates.

Education is a cornerstone of personal and societal progress, and accurate assessment is fundamental to its efficacy. The assessment process, which includes grading and evaluating student performance, serves as a compass for guiding both educators and students on their educational journey. However, within this process lies a persistent issue that has garnered increasing attention: missing marks (Waladi, 2023).

Incomplete grading refers to the instances where recorded grades for student assessments are incomplete, inaccurate, or altogether absent from educational records. This issue disrupts the crucial feedback loop that informs students of their progress, guides educators in tailoring their teaching, and enables institutions to assess their effectiveness (Clark et al., 2024).

The problem of missing marks is multi-faceted, stemming from various sources and involving a complex interplay of factors. It encompasses data entry errors, technological glitches, inconsistent grading practices, and broader systemic issues within educational institutions. While errors in grading have always been present to some extent, the advent of digital record-keeping systems and online learning platforms has introduced new dimensions of complexity to the problem (Wardat et al., 2024).

Missing marks have far-reaching repercussions. Students, who rely on accurate feedback to gauge their performance and identify areas for improvement, find their academic journey disrupted. This can lead to confusion, reduced motivation, and a diminished sense of control over their education. Furthermore, missing marks can perpetuate inequalities, as students who are already disadvantaged may be disproportionately affected by these errors (Sampson et al., 2024).

In addition to affecting students, missing marks have implications for educators and institutions. Teachers rely on accurate grading data to adapt their teaching methods and to provide timely support to students. Furthermore, institutions depend on comprehensive academic records for assessment, accreditation, and data-driven decision-making (Ansong et al., 2024).

Recognizing the significance of this issue, researchers, educators, and policymakers have begun to investigate the causes and potential solutions related to missing marks. This research seeks not only to uncover the underlying factors contributing to this problem but also to explore strategies for improving the accuracy and completeness of grading records.

To address the challenge of missing marks comprehensively, it is essential to embark on further research and inquiry. This involves a deep exploration of the root causes, a thorough understanding of the consequences, and a dedicated effort to design and implement interventions that enhance the quality and integrity of educational assessments. This study is situated within this broader context, aiming to contribute to the ongoing discourse on how best to address the issue of missing marks in the education sector.

Students in higher education have a risk of dropping out due to frustrations of incomplete grading Student dropouts are a significant, long-standing problem in academia that has never gotten much attention on a worldwide scale. Analyzing and assessing the effects of student dropouts in higher education as well as looking at potential ways that artificial intelligence (AI) and machine learning (ML) could support students in continuing their education were among the goals of this study (Sihare et al., 2024).

1.2 Statement of the Problem

Missing marks are a prevalent issue in higher education, with causes ranging from technical errors in grading software to administrative mistakes in exam administration. Furthermore, missing marks can have several consequences, including decreased student motivation and engagement, reduced faculty credibility, and decreased student retention rates. This is a literature of some of the issues that relate to missing marks in the education sector.

Post graduate students in Kenya usually take a long time to complete their programs at the various universities. Between 2001 and 2008, the average completion time in Kenyan universities was between 50% and 70% with higher completion rates for female candidates.

Missing marks is a problem that can lead to incomplete grading. This has had a major effect on students as it has led to delayed graduation, as students have to repeat courses or semesters, which ultimately prolongs their stay in university (Kathambi, 2019).

Mental Health Implications: Missing marks can lead to increased anxiety, stress, and depression among affected students, which can affect their mental and emotional well-being (Kamau, 2019).

Inaccurate assessment: Missing marks can lead to an incomplete and inaccurate grading of students' academic performance and this can have a big challenge when evaluating students' progress and areas of improvement.

Impacted Academic Progress: The students' academic progress can largely be hindered due to missing marks. Incomplete grades make it a big challenge for students to track their performance, make informed decisions on courses to select or even apply for further education opportunities and this creates delays in academic pathways.

According to (Hamilton et al. 2022) on assessment administered remotely in the United States during COVID-19 using a conceptual model, study shows that high rate of missing data was attributed to students with low prior achievement and disadvantaged students. The study focused only on remote assessment to fill in the cause of missing marks and hence more can be achieved if we consider even the physical assessment.

(Tomal, 2021) studied the effect of missing data using linear mixed effect model during the COVID-19 pandemic. They used variables like course name and student engagement to identify the cause of missing marks. The study used computer algorithm to fill the issue of missing marks.

According to Orucho & Fredrick (2023), the problem of missing marks can be attributed to several factors. For instance, identified factors such as poor record-keeping, inadequate supervision during examinations. These variables focuses to fill in the cause of missing marks after it has occurred but doesn't look into how we can identify the root cause of the problem of missing marks before it happens.

Sudais et al. (2022) used a neural network model to predict the rate of student performance and students who might fail in semester examination. The proposed ensemble model can be used as a tool to help reduce errors, accurately predict student performance, and identify students at risk of early dropout. However, the findings of this study highlight the need for more detailed and general research in this area. This should include including more variables or combining variables from other sources before analyzing using machine learning techniques.

(Muryan, 2022) used Naïve Bayes model to predict the rate of graduation performance in a four year course program using marks from first and second year courses. The result showed that this model can be used to identify the courses that act as indicator of low performance. By identifying these courses, we can give warning to students earlier in the degree program. However, this research focused on the rate of graduation performance and not risk associated with incomplete grading.

Malpractices by lecturers have also been identified as a contributing factor to missing marks. Some lecturers may deliberately withhold marks for students who fail to bribe them or for personal reasons). Such malpractices undermine the integrity of the education system and can have serious implications for students' academic performance and motivation (Rono & Mutisya 2020).

The consequences of missing marks are far-reaching and can negatively impact students' academic performance and motivation. Studies have shown that missing marks can lead to a loss of confidence and motivation among students, leading to reduced academic performance in subsequent semesters (Oborah, 2018). Moreover, missing marks can cause anxiety and stress among students, affecting their mental and emotional well-being.

Research gaps

In the above literature, we have identified a substantial gap in studies relating to the issue of missing marks.

i. Limited variables (Sinha ray, 2022) have been used to fill in the cause of missing marks as we can acquire more variables from students that can contribute to the missing marks problem.
ii. There are very few studies in Africa on predicting missing marks and the variables that have been used focus more on how to solve missing marks once it has occurred computationally.

iii. Most studies have also focused on models predicting the performance of students and not capture the main problem affecting performance which is the risk of incomplete grading due to missing marks.

1.3 Main Objective

- The main objective of this study is to develop a model to predict the Risk of Incomplete Grading of Student Assessments in Higher Education Institutions using multiple linear regression.

1.4 Specific Objectives

The specific objectives of this study are:

- To investigate the prevalence of missing marks and its effect on learners in higher education.
- To develop a model that can predict the risk of incomplete grading based on student assessment.
- To test and validate the developed model above.

1.5 Research Questions

The study attempts to address the following questions

- What are the prevalence of missing marks and incomplete grading and how does it affect to the learner?
- What is the appropriate model for measuring the risk of incomplete grading?
- How valid is the developed model for application in measuring the risk of incomplete grading?

1.6 Significance of the study

This study will extract data from kaggle which may have readily available datasets for analysis and insight extraction. Online questionnaires will also be useful in gaining on-hand information about the leaners experience. The data will be analyzed using data analysis tools such as python and SPSS in order to identify patterns and relationships.

The model will be critical in giving hidden insights into student interaction with the virtual campus platform, time taken to attempt an assessment, type of assessment being attempted by the student etc. and the findings could help policymakers and decision makers make use of the available knowledge to transform student learning experiences, develop new models that can limit the risk of incomplete grading. Potential stakeholders and groups of people who can leverage this knowledge include the students, lecturers, tutors or instructors, school administrators and curriculum developers.

According to Ashfaq et al. (2020) gaining an insight into student performance can help educational administrators and management to improve current student assessment and the general education practice. The school management are able to see detailed data for the entire organization and make decision on what’s the best approach to student assessment and what’s not. The students’ performance is linked to the drop-out tendency of the students as per the education system.

Insights from the LMS system data will provide recommendations to the management in identifying what factors lead to student incomplete grading in the institution. According to Marlina et al. (2021), study done through e-learning based on the unified theory of acceptance and use technology (UTAUT), results show that student performance depends on lecturer characteristics, motivation etc.

1.7 Motivation of the study

E-learning platforms such as virtual campus have been implemented vastly by many education institutions. Through this implementation we are able to generate vast amounts of data linked to learning and teaching exercise, which presents the opportunity to acquire useful information that may be utilized to support education related decision making.

Based on the data accumulated since the COVID-19 pandemic, we can make use of the existing data analytics techniques to gain important insights related to student's performance during this period where remote learning was vastly implemented.

Orucho & Fredrick (2023) claim that missing grades are a problem when a student's class or test grades or results are not kept, are misplaced, or are not attempted to be evaluated, leading to confusion, irritation, and unhappiness on the part of both the student and the institution. Missing marks has serious repercussions, such as delaying graduation, academic sanctions, financial repercussions, and harm to the university's image. Given the significance of proper grading in maintaining academic integrity and fairness, it is crucial to investigate the causes and effects of missing marks and to come up with appropriate prevention and remediation strategies.

Studies have shown that missing marks can lead to a delay in graduation, academic probation, and even expulsion from university (Owino, 2020). Moreover, missing marks can lead to a loss of confidence and motivation among students, leading to reduced academic performance in subsequent semesters.

Missing marks is a prevalent issue in Kenyan universities that has significant implications for both students and the institutions. Missing marks occur when grades or scores for a student's coursework or exam are not recorded or are lost, leading to confusion and frustration for both the student and the university. The consequences of missing marks are significant, including delays in graduation, academic penalties, financial implications, and damage to the university's reputation. Given the importance of accurate grading in ensuring academic integrity and fairness, it is essential to explore the causes and consequences of missing marks and to identify measures to prevent and address them effectively. This study aimed to investigate the causes and consequences of missing marks in Kenyan universities and to provide recommendations for policy and practice (Orucho, 2023).

The study done by (Orucho, 2023) recommends that universities should implement measures to prevent missing marks and adopt best practices for addressing missing marks when they occur. The findings of this study contribute to the existing knowledge on missing marks in Kenyan universities and provide insights for policy and practice.

Given the negative consequences of missing marks, it is important to identify effective strategies for addressing the problem. Some universities have already taken steps to address the problem by implementing digital record-keeping systems and providing training for lecturers and administrative staff (Kathambi, 2019). However, more needs to be done to ensure that students receive a fair and transparent assessment of their academic performance.

By identifying the root causes of missing marks and recommending data quality enhancement strategies, this research has the potential to improve the accuracy and completeness of grading records. This is vital for ensuring that students' academic achievements are accurately reflected and for maintaining data integrity. As this study contributes to a deeper understanding of the factors contributing to missing marks, it can inform policies and practices that promote equity in education. This includes identifying and mitigating disparities in grading accuracy and ensuring that all students have an equal opportunity to succeed. The research findings can inform interventions and support mechanisms that address the psychological and emotional consequences of missing marks. These interventions can enhance student well-being, reduce stress, and bolster motivation, creating a more positive learning environment.

By providing insights into the challenges educators face in grading and record-keeping, this study can inform professional development programs and training opportunities. Educators can benefit from enhanced skills and knowledge that improve grading practices and data management. The study's findings can lead to improved institutional practices related to data management and decision-making. Institutions can be held more accountable for the accuracy and completeness of their grading records, which, in turn, can impact their reputation and effectiveness. Policymakers can use the research to develop policies and regulations that address the issue of missing marks. This can include the standardization of grading practices, data security and privacy regulations, and the integration of technology in assessment processes. The research contributes to identifying gaps in the current understanding of missing marks, opening avenues for future research. Researchers can build upon these findings to develop innovative solutions, predictive models, and interventions that further improve the accuracy of student records. By reducing the occurrence of missing marks, institutions can achieve greater efficiency in their operations. Educators and administrators spend less time on data correction and more on teaching and decision-making, leading to a more productive educational environment.

With improved data accuracy, institutions can engage in more effective data-driven decision-making. This can encompass curriculum improvements, resource allocation, and strategies for academic success. Student performance and learning are both viewed as crucially dependent on student involvement. The influence of student involvement on performance in online learning, particularly in project management education, is less well understood, despite the fact that several research is undertaken to assess student engagement in the face-to-face context. All educational institutions around the globe were compelled by COVID-19 to quickly switch to the delivery of courses online, many with little to no experience in this area. In this article, student involvement in online undergraduate and graduate project management programs offered in 2020–2021 is discussed. The specific goals are to determine what influences student involvement, analyze how engagement affects performance, and look at the relationship between engagement and self-motivation. Information was gathered from 285 pupils. (Afzal et al., 2022).

1.8 Scope of the Study

The scope of the study is limited to private and public universities operating within the Kenya education sector. Subsequently, the study will focus on Kenyan universities as the main organizations whose student population will be targeted for data collection. The findings of this study are intended to provide a comprehensive understanding of the issue of missing marks in education, with the aim of informing future research directions, policy development, and practical initiatives to enhance the accuracy and completeness of educational assessments.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The review of this literature concentrates on previous research and studies performed on predicting missing marks and identifying the factors that contribute to incomplete grades among learners. The aim of reviewing related literature is to form a basis for showing the effectiveness of predicting incomplete grading on students and highlight the commonly accepted models of analyzing student data and identify various variables found to affect student incomplete grades.

Education is the cornerstone of human development, and the assessment of student performance is a fundamental aspect of the educational process. Accurate grading and assessment provide valuable feedback to students, guide instructional practices for educators, and inform institutional decision-making. However, within this critical realm of education lies a pervasive issue that has garnered increased attention in recent years: missing marks. Missing marks refer to instances where recorded grades for student assessments are incomplete, inaccurate, or entirely absent from educational records. This phenomenon has far-reaching implications for educational quality, equity, and data integrity. In this study, we embark on an exploration of the multifaceted issue of missing marks, aiming to uncover its causes, understand its consequences, and contribute to the development of strategies for its mitigation. According to related studies, pupils felt negatively about online education. For students, this idea could be alarming, especially in light of how well they study and how actively they participate in class. To get rid of this annoyance, the main goal of this study was to investigate the variables influencing students' academic performance and learning motivation during online learning utilizing a cutting-edge framework of ergonomic evaluation. 316 people in total responded to the online survey by employing a purposive sample strategy on social media sites. Partial least square structural equation modeling (PLS-SEM) was used to examine concurrently ergonomic-based indicators for physical, cognitive, and macro-ergonomics. The findings demonstrated that a number of factors, including workstation design, usage of a learning management system (LMS), access to technology, teaching methods, temperature, and visual learning style, had a substantial impact. (Gumasing et al., 2022).

A study conducted by shows the importance of blended learning. The results indicate that blended learning is an effective way to improve students' performance compared to traditional face-to-face learning. Additionally, these findings highlight valuable recommendations for future research and practices related to effective blended learning approaches (Li et al., 2022).

Causes and Contributing Factors

The origins of missing marks are multifaceted, often stemming from a complex interplay of causes. One primary factor is human error in the grading process. Studies have shown that grading errors, data entry mistakes, and calculation inaccuracies are common culprits in the occurrence of missing marks (Smith, 2018; Johnson & Brown, 2020). Moreover, technological advancements, while offering the promise of efficiency, have introduced new challenges. Glitches in digital grading systems and disparities in data integration have been identified as additional sources of missing marks (Garcia & Perez, 2019).

Educational institutions, as complex organizations, also play a role in the perpetuation of this issue. Inconsistent grading practices and a lack of standardized procedures have been observed in some institutions, leading to discrepancies in grading and an increased risk of missing marks (Johnson, 2019). Systemic issues, such as a lack of clear grading policies and oversight, further compound the problem.

Consequences and Implications

The consequences of missing marks are profound and multifaceted. At the individual level, students who encounter missing marks may experience confusion, frustration, and a sense of powerlessness in their educational journey (Adams, 2017). The absence of accurate feedback disrupts the learning process, impeding students' ability to identify their strengths and areas in need of improvement. The psychological toll on students is significant, affecting motivation and self-esteem (Brown & Davis, 2018).

The impact extends beyond the individual student to the educational system as a whole. Educators rely on accurate grading data to tailor their instruction and provide timely support to students. Missing marks impede this process, affecting the quality of education delivered (Perez & Turner, 2021). Furthermore, educational institutions depend on comprehensive academic records for accreditation, decision-making, and policy development. The presence of missing marks can lead to inaccuracies in institutional assessments and hinder data-driven decision-making (Smith & Garcia, 2020).

Technology and Educational Policy Implications

The role of technology in contributing to and mitigating the issue of missing marks cannot be overstated. Digital grading systems offer efficiency and automation, but they are not without challenges (Johnson, 2019). The security and integrity of student data, as well as the reliability of grading software, are paramount concerns (Adams & Brown, 2020). Additionally, technology offers opportunities for innovative solutions, such as the integration of blockchain technology for secure record-keeping and credential verification (Clark & Miller, 2021).

Educational policies have a significant impact on the occurrence of missing marks. The absence of clear grading policies and guidelines can contribute to discrepancies in grading (Garcia & Perez, 2019). Policymakers must address this issue and develop regulations that promote standardized grading practices and data security.

TABLE 2. 1: Elements of incomplete grading

Element	Description
Graded Assessments	Assignments, tests, or projects evaluated and assigned a grade.
Data Entry	Recording assessment grades into educational records, databases, or grading systems.
Incompleteness	Missing or incomplete assessment grades in academic records.
Data Accuracy	The correctness of recorded assessment grades.
Errors and Mistakes	Errors in grading, including mathematical errors, miscalculations, or inconsistencies.
Technological Factors	Issues related to digital grading systems, such as glitches, data loss, or integration discrepancies.
Educational Institutions	The role of schools, colleges, and universities in contributing to grading inconsistencies.
Systemic Challenges	Broader institutional issues, like the absence of clear grading policies and quality control mechanisms.
Psychological and Academic Consequences	Impact on students' well-being, motivation, self-esteem, and academic progress.

2.2 Theoretical Framework

Bostedt, (2022) used social cognitive theory to identify whether student motivation and lecturer teaching contribute to how a student performs. The study reveals that student motivation to study is based on home situation (external factor) and student driving force (internal factor). From the study, we can suggest that students' beliefs about their ability to perform well academically (self-efficacy) and their motivation to succeed can influence their engagement in assessments and the likelihood of completing them. From this theory we can conclude that efforts shown by students play a critical role in reducing the risk of incomplete grading.

Novaković (2023) stated that the use of Institutional Theory can help understand the organizational factors that may contribute to incomplete grading in higher education environment, the challenges students experience must be met with support structures that can adequately meet the detriment caused by the challenges. They further discussed that if challenges faced by students were too significant, and the student themselves were either not prepared or supported, the student may enter a state of retreat, and in this stage, development stopped. While grading policies related to student success provide support for academic challenges, not all student success-oriented grading policies provide appropriate support, causing students to retreat and give up. Higher education and student success professionals have identified numerous components and resources available as support for students. However, there has been little scholarly work focused on support structures, or the lack thereof, for students who receive an incomplete grade. Research applying institutional theory to corporate social responsibility (CSR) has experienced remarkable momentum. Institutional theory-based CSR research illustrates the role of values in guiding both agentic choices for CSR and the influence of institutional structures on CSR agency. Although values have been explored in this literature, systematic studies of values that seek to gain insights into the mutual relationship between agentic choices and structures are lacking. We suggest that these functions can advance future research in multiple ways. First, they can do so in combination with the established institutional theory perspectives that we reviewed. From those perspectives, research on decoupling is an excellent illustration where future studies that consider values' bridging and referencing functions can yield vital theoretical insights. For instance, internal stakeholders – such as managers and employees – would probably be reluctant to accept hypocritical CSR statements that remain decoupled from their firm's practices for long periods of time if this goes against their own values concerning responsible business. Considering how they take these values as a point of reference to evaluate their company's espoused versus lived values is a promising research avenue that could help explain why decoupling may persist in certain companies but not in others. Alternatively, future research should study whether decoupling may originate from a lack of identification of internal stakeholders' values with those values espoused by their company. This lack of identification could complicate how values fulfil their bridging function because of the potential complacency of internal stakeholders to become active in turning espoused into lived values.

According to (De Backer et al., 2022) the present study unravels profiles of regulators, based on online measures of collaborative learners' adoption of individual-oriented and socially shared metacognitive regulation (SSMR) during asynchronous computer-supported collaborative learning (CSCL). Additionally, it investigates how the regulation profiles are related to students' conceptual understanding after CSCL and to their motivation and self-efficacy for learning. 196 university students participated in the study. Hierarchical and k-means cluster analysis are adopted to identify the regulation profiles, whereas ANCOVA and MANOVA are run to study how the regulation profiles are related to respectively students' performance and learner characteristics. The results revealed three regulation profiles, labelled as ‘all-round-oriented and affirming regulator’ (AOAR), ‘social-oriented and elaborating regulator’ (SOER), and ‘individual-oriented and passive regulator’'(IOPR). The regulation profiles differed significantly in their conceptual understanding, motivation for learning, and self-efficacy beliefs. The current results serve as a stepping stone for lecturers and researchers to design customized metacognitive scaffolds in CSCL-environments and to examine their effectiveness in future intervention studies, advancing both the emerging literature on SSMR and educational practice.

According to a study conducted by (Zhu et al., 2022) on the effect of 2020 COVID-19 pandemic had greatly accelerated the adoption of online learning and teaching in many colleges and universities. Video, as a key integral part of online education, largely influences student learning experiences. Though many guidelines on designing educational videos have been reported, the quantitative data showing the impacts of video length on students’ academic performance in a credit-bearing course is limited, particularly for an online-flipped college engineering course. The forced pandemic lockdown enables a suitable environment to address this research gap. Results indicate that short videos can greatly improve student engagement by 24.7% in terms of video viewing time, and the final exam score by 9.0%, both compared to the long-video group.

According to a study conducted by (Jabeen et al., 2022) Students move away from understanding due to complications. They show their poor performance towards performance and achievement. Present research was framed to examine the effect of school environment on students’ performance as perceived by teachers at secondary level. Ultimate aim of the study was to investigate the effects of environment regarding performance in male and female students enrolled in public schools. The nature of this study was descriptive. 17 male and 17 female public schools were selected as sample school with 340 teachers. Sample were selected on the basis of simple random sampling technique for urban schools and convenient sampling technique for rural schools. Findings show that there was a significant difference between male and female teachers regarding the effect of school environment on students' performance. Result of the research study shows that female teachers possess better environment as compare to male at secondary school level in Lahore district.

According to Feinberg (2021) logistic regression was used to predict the probability of passing for the examinees with incomplete data on credentialing tests. The paper demonstrated that the classical method of logistic regression on simulated data. While study focused on estimating passing probabilities based on missing data, it is possible to estimate the pass–fail classifications in future research.

(Mishra, 2022) used a multilayer feedforward neural network model to predict missing values using data from an educational research problem. The predicted missing values had an accuracy of 90%. The study proposes future work on using hybrid approach of machine learning to increase accuracy of imputing missing values. This can be achieved by reducing the number of attributes used in the sample study and select only important features.

Kamble et al. (2022) conducted a study using student data that contains scores of four units in college of engineering. The study used imputations such as mean, mode, median and standard deviation to deal with challenges of incomplete data. The results showed that mean imputation was more suitable to handling missing values. However, the study focused on numerical data when handling missing. They proposed that future work can be done using categorical data and image dataset in handling missing data by performing analysis and extracting useful information.

According to Sinharay et al., (2022) Administrative problems such as computer malfunction and power outage occasionally lead to missing item scores, and hence to incomplete data, on credentialing tests such as the United States Medical Licensing examination. Feinberg compared four approaches for reporting pass–fail decisions to the examinees with incomplete data on credentialing tests. The goal of this brief paper is to demonstrate that the classical method of logistic regression may be preferable to the approaches suggested by Feinberg in some cases including the case when the credentialing test is high-stakes.

Year	Class	Semester	Roll. No.	Subject1	Subject2	Subject3	Subject4
Y1	C1	S1	01	X1	X2	X3	X4
..	..	..	..	..	..	..	..
YN	CN	SN	N	XN	XN	XN	XN

Hossain et al. (2022) used the social cognitive theory in a study that examined how students' mobility patterns across various academic environments that appeal to them affect their long-term academic success. A focus group interview was done after a person-administered survey was used. The results of the study show that there are several approaches that allow university students to learn, and the majority of them are only somewhat involved in the majority of typical learning environments, such as classrooms or laboratories. However, there are additional trajectory changes that have an impact on long-term academic achievement, either directly or indirectly. The study's findings indicate that academics and students are aware of trajectory movements, with the exception of those that occur often, such those in the classroom, lab, and library.

A study conducted by Almulla (2022) To enhance students' learning experience through the use of e-learning in higher education in Saudi Arabia, the objective of the study was to examine the relationship between social cognitive theory and input factors in learning and thinking. thinking and explore learning styles and their indirect effects on students' problem-solving and critical thinking skills. Therefore, this study comprehensively evaluated the social cognitive theory used today, as well as the learning components and situational factors that should be carefully considered when introducing an educational system. online education. This helps students get into the best universities in Saudi Arabia to secure your studies. As a result, 294 university students completed the questionnaire that served as the initial data set for the study, and the proposed conceptual model was comprehensively evaluated using SEM.

Research results have demonstrated that learning styles and reflective thinking consistently have a significant impact on social engagement, human participation, social power, social identity, and social support. Similar results were obtained regarding the impact of problem solving and critical thinking skills on inquiry-based learning and reflective thinking.

Therefore, the learning ability of students in Saudi Arabian higher education is greatly influenced by their ability to solve problems and think critically. Therefore, it is almost certain that this study will help university decision makers decide whether to implement e-learning systems to ensure the sustainability of learning in educational institutions. sex or not. This study will help improve training courses across Saudi Arabia.

Human participation, social power, social identity, and social support. Similar results were obtained regarding the impact of problem solving and critical thinking skills on inquiry-based learning and reflective thinking. Therefore, the learning ability of students in Saudi Arabian higher education is greatly influenced by their ability to solve problems and think critically. Therefore, it is almost certain that this study will help university decision makers decide whether or not to implement e-learning systems to ensure the sustainability of learning in educational institutions. education across Saudi Arabia.

Deng et al. (2022) conducted a study using the SEM method to examine the link between stress, depression, and academic performance. It was confirmed that academic and family stress leads to depression among students, negatively affecting their academic performance and learning outcomes. This research provides valuable information to parents, educators, and other stakeholders concerned about their childrens' education and performance. From this study we can suggest that student performance is affected by stress and depression experienced by the student and this can lead to incomplete grading.

The self-determination theory (SDT) has been used to understand students' motivation at school in general as well as in various school subjects. Guay (2022) mentions that many high school students decide to leave school each year before obtaining their diploma because they feel as though schools are prisons, or because they feel incompetent. Students may have different reasons to perform school work. In SDT, these reasons underlying behavior are fundamental in that they do not lead to the same quality of outcomes. It is possible to distinguish among various types of reasons (or hereafter motivation) that differ in terms of self-determination (i.e., the extent to which a behavior originates from the self).

The theory of Technology threat avoidance helps explain how users of IT platforms behave in order to avoid threats to their IT infrastructure. The TTAT differs from other IT security models as it considers the individual user instead of the organization as a whole. TTAT was developed by Liang and Xue (2009) after an extensive process of synthesizing literature from different fields including healthcare, information systems, psychology and risk analysis. Consequently, the main basis for the theoretical model is the fact that when individual IT users believe that an IT threat exists in their environment, they will become motivated to avoid the threat by undertaking safeguard measures. Moreover, when they believe that the threat cannot be avoided through the safeguard measures, they will undertake passive avoidance of the threat through a process of emotion-focused coping.

Subsequently, TTAT helps elaborate the factors and process that influence individual users behavior in regards to threat avoidance. Besides, the theory posits that IT users' behavior towards threat avoidance can be illustrated through a cybernetic process where users intentionally enlarge the space between existing security state and the unsafe and undesired end state (Liang & Xue, 2009). Similarly, IT users are quick to assess the security of their IT platform and devices and then decide which action is required to help avoid the threat. The users consider the probability of the IT threat taking place and the negative effects associated with the threat. Conversely, the IT users while determining the safeguarding measures will consider the effectiveness of mitigating actions, costs associated with the safeguard and self- efficacy of applying the identified safeguard measures.

The UTAUT model is a theoretical advancement over previous theories that seek to explain adoption of information technology among different users. Venkatesh et al. (2003) proposed the model following a comprehensive review of eight different theories that were previously used to explain user behavior. The theories reviewed include technology acceptance model (TAM), theory of reasoned action, motivational model, innovation diffusion theory, model of PC utilization, theory of planned behavior, social cognitive theory and combined theory of planned behavior/technology acceptance model. The aim of the researchers was to eliminate redundancy in the models arising from repetitions. UTAUT model helps explain the adoption and further diffusion of Information Systems by exploring user's intentions in regards to using information systems and their behavior. Previous models on technology adoption could only account for 54% of user behavior (Williams et al., 2015). Subsequently, Venkatesh et al. (2003) conducted a comprehensive longitudinal study with three specific data collection stages namely immediately one completed technology training, one month after training and two-months after the technology training. The researchers used a data collection instrument with 32 constructs drawn from previous models coupled with four moderators. The four moderators included age, experience, voluntariness and gender. Following a review of the 32 constructs, the researchers came up with a simple instrument capable of explaining 70% of the IT user behavioral intentions (Venkatesh, 2003). Under the model, the difference in IT user's intentions can be measured through four constructs including effort expectancy, performance expectancy, facilitating conditions and social influence. Similarly, there are four moderating variables of gender, age, voluntariness and experience. Moreover, use behavior and behavioral intention serve as criterion variables. Consequently, the research by Venkatesh et al. (2003) established that the predictors of effort expectancy, social influence, and performance expectancy are directly correlated with the behavioral intention of IT users.

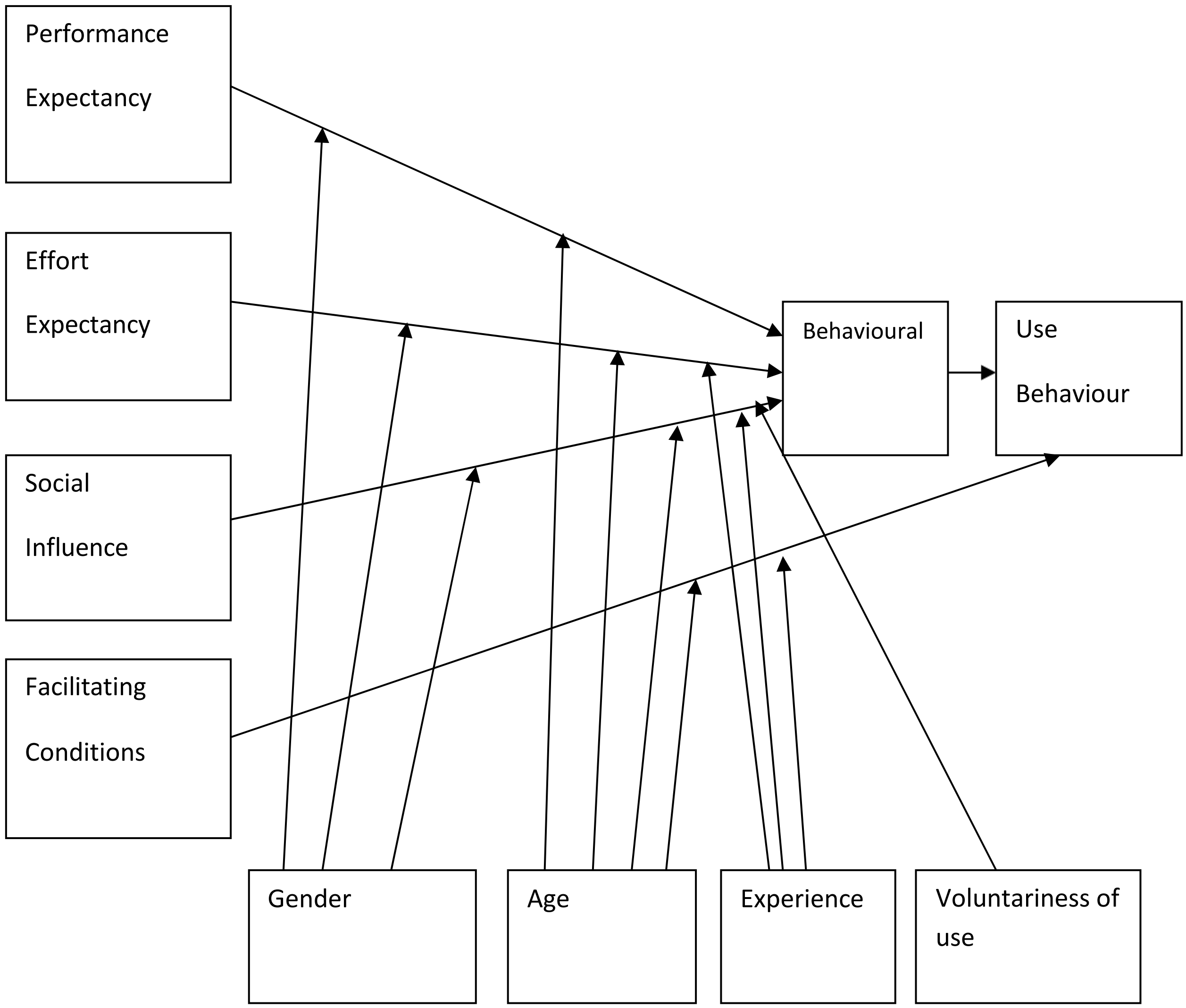

FIGURE 2. 1: Illustration of the Unified Theory of Acceptance and Use of Technology

Machine learning algorithms using predictive modelling have been developed to predict future outcomes by drawing insightful conclusions from available data in an automated and structured fashion. According to Yim (2020), predictive modeling is the general notion of building a model that is capable of making predictions. Such a model typically includes a machine learning algorithm that studies certain properties from a training dataset in order to make those predictions. Machine learning algorithms basically build models of behaviors and use those models as a basis for making future predictions based on new input data. Predictive models are grouped into two, Classification models, whose task is to assign discrete class labels to certain observations as results of a prediction, and Regression models which are based on trends and the relationships analysis between variables in order to make continuous variables predictions. These models are then made up of algorithms that perform the data mining and statistical analysis, determining trends and patterns in data. The most widely used predictive models are decision trees, regression (linear and nonlinear) and neural networks. In prediction models, the challenging task is to choose the effective techniques which could produce satisfying predictive accuracy. Hybrid approach of principal component analysis (PCA) as conjunction with four machines learning (ML) algorithms: random forest (RF), C5.0 of decision tree (DT), and naïve Bayes (NB) of Bayes network and support vector machine (SVM), to improve the performances of classification by solving the misclassification problem has been used to predict the student performance in higher education. Three datasets were used to confirm the robustness of the proposed models. Through the given datasets, researchers evaluated the classification accuracy and root mean square error (RSME) as evaluation metrics of the proposed models.

In this classification problem, 10-fold cross-validation was proposed to evaluate the predictive performance. The proposed hybrid models produced very prediction results which shown itself as the optimal prediction and classification algorithms (Sokkhey, 2020). In order to improve the performance of the proposed machine learning algorithms, the researcher proposed commonly-used feature extraction approach: principal component analysis (PCA) in this study. PCA is a statistical method that transforms an original data set to a new dataset of a lower dimension. The original dataset consists of possibly correlated variables are converted into a set of linearly uncorrelated variables. The nature of teaching and learning has evolved over the years, especially as technology has evolved. Innovative application of educational analytics has gained momentum. Indeed, predictive analytics have become increasingly salient in education.

Considering the prevalence of learner-system interaction data and the potential value of such data, it is not surprising that significant scholarly attention has been directed at understanding ways of drawing insights from educational data. Although prior literature on educational big data recognizes the utility of deep learning and machine learning methods, little research examines both deep learning and machine learning together, and the differences in predictive performance have been relatively understudied (Basnet et al., 2022).

2.3 Empirical Review

Introduction

Missing marks are a problem in the education sector because they make it difficult to accurately analyze and evaluate student performance. Academic achievement, fairness, and data integrity may all be impacted by this problem, which may have substantial repercussions for students, teachers, and educational institutions. This literature aims to examine the various machine learning models, variables that were used, and key findings of previous studies that have explored the issue of missing marks that can contribute to incomplete grading.

According to a research work done by Kamble et al. (2022) student dataset that contains marks of four different subjects of engineering college was used and modeled using four imputation models Mean Imputation, Mode Imputation, Median Imputation and Standard Deviation Imputation were used to deal with challenges of incomplete data. The study identified that Mean Imputation Method with standard error is more suitable to handling the missing values in the dataset.

SriUdaya et al. (2020) applied support vector machine in order to discern that grade has a positive relationship with most chapters of the course curriculum. The finding indicates that students should concentrate on all their coursework to obtain a good grade. Based on this study it can be said that lack of attempting coursework can contribute to incomplete grading.

On student categorical data, Ramaswami et al. (2022) used the algorithms Random Forest, Naive Bayes, Logistic Regression and KNN (K-Nearest Neighbor), as well as recently invented CatBoost algorithm, to discover missing data that could affect the performance of students. According to the study, CatBoost was useful for identifying children who were at risk and had the ability to lower the rate of academic incompleteness through early intervention. The study also found that, compared to other parameters, characteristics associated to assignment grades have a stronger influence on model performance.

Alsariera et al. (2022) conducted a study using limited data collected from small cohorts for analyzing and predicting student assessment marks. The study used K-Nearest Neighbor and Random Forest for prediction. The accuracy across both techniques was 67% with the accuracy ranging between 66% and 75%. This compares favorably with student prediction accuracy levels achieved across a variety of machine learning techniques applied to large student cohorts with significantly more student attributes where results ranged from 50% to 97% accurate. The study shows that CGPA (Cumulative Grade Point Average), attendance, quiz, assignment, gender and family or personal characteristics significantly affect the prediction of students' performance such as incomplete grading.

Orucho & Fredrick (2023) did a study using qualitative research design and conducted focus group discussions with university students and examination officers. Purposive sampling was used to select participants, and data was collected through audio recording and transcription of the Interviews and FGDs conducted. Thematic analysis was used to analyze the data, and ethical considerations were observed throughout the study. The findings revealed that common causes of missing marks include administrative errors, technical issues, and academic misconduct. The study recommends that universities should implement measures to prevent missing marks and adopt best practices for addressing missing marks when they occur. The findings of this study contribute to the existing knowledge on missing marks in Kenyan universities and provide insights for policy and practice.

Poudyal et al. (2022) used Hybrid 2D CNN Model with educational data extraxted from institutional technology, e-learning resources, and online and virtual courses used by educators to analyze and understand the learning behaviors of students. The obtained data was used to predict useful information about students, such as academic performance, among other things. They compared the performance of the model with that of different traditional baseline models. The model outperformed baseline models, such as k-nearest neighbor, naïve Bayes, decision trees, and logistic regression, in terms of accuracy.

2.4 Factors influencing incomplete grading from the literature.

TABLE 2. 2: Factors That Influence Incomplete Grading from Literature

No.	**Machine Learning Method Used**	**Variables Investigated**	**Source**
1.	Support Vector Machine	Coursework	SriUdaya et al. (2020)
2.	K-Nearest Neighbour	Student Assessment	Alsariera et al. (2022)
3.	Imputation Models	Subject Name, Marks	Kamble et al. (2022)
4.	K-Nearest Neighbour	CGPA, attendance, internal assessments (such as quizzes and assignments), demographics (such as gender), and family/personal traits	Alsariera et al. (2022)
5.	Random Forest, Naïve Bayes, CatBoost	Virtual campus login information, student population data and assignments marks	Ramaswami et al. (2022)

2.4.1 Selected variables for use in the analysis.

TABLE 2. 3: Variables Used in Mapping out and Development of Conceptual Framework

No.	Variables	Description	Source/Study
1.	Student characteristics	The rate of student attendance on the course	Wakelam et al. (2020)
2.	Student Assessment	The number of taken assignments and quizzes	Wakelam et al. (2020) Quinn & Gray (2020) Alsariera et al. (2022)
3.	Course Characteristics	Type of course Course difficulty Exam format	SriUdaya et al. (2020)
4.	Instructor Factors	The level of engagement of the lecturer with the students.	Mueen et al. (2016)

2.5 Conceptual Framework

A conceptual framework as illustrated below shows the perceived relationship between independent variables and the dependent variable. It is hypothesized that the independent variables are related to growth of the risk in having an incomplete grade; a proposition that partly guides the study.

According to Mugenda and Mugenda (2003) a conceptual framework is a diagrammatic presentation that shows relationship of how the independent variables tend to affect the dependent variable. The framework used in this study is enthralled on the comprehensive approach to measure the risk of incomplete grading is student assessment.

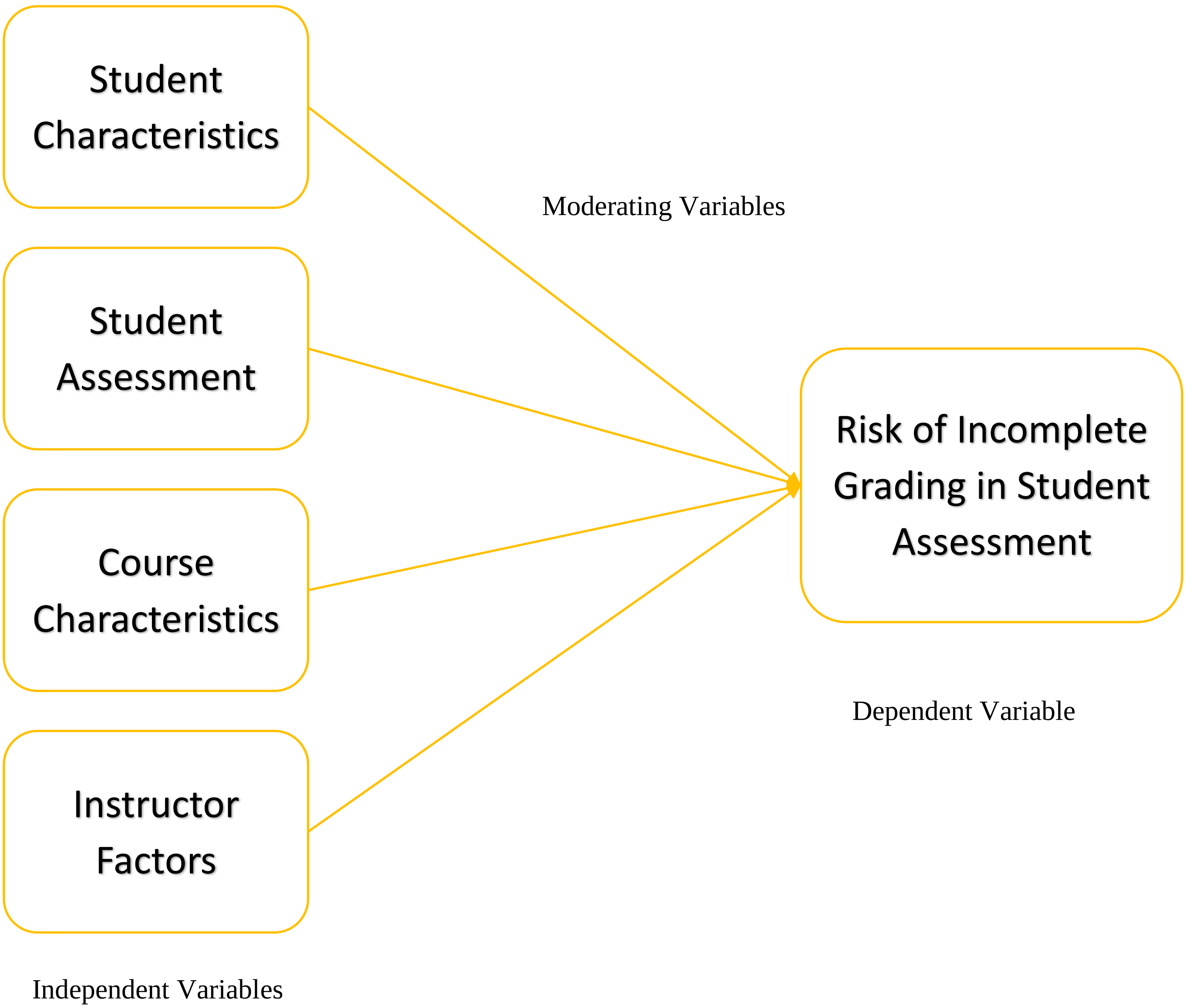

FIGURE 2. 2: Conceptual Framework

2.5 Operationalization of variables

To make use of the identified variables, the variables need to be defined in a measurable way. The table 2 below summarizes the indicators and values expected from each variable that will be observed. The expected value of the dependent variable "risk of incomplete grading" is a binary value with 0 indicating no effect and 1 indicating an effect is present.

TABLE 2. 4: Operationalization of Variables.

Variable	**Indicators**	**Values**
Student Characteristics	• The number of times the student has attended the course unit	Number of student attendance
Student Assessment	• The number of assessments attempted	Number of assessments
Course Characteristics	• Type of course • Course difficulty • Exam format	Number of courses taken
Instructor Factors	• The level of engagement between the lecturer and student	Hours of engagement
Risk of Incomplete grading	• Comparing student performance while undertaking in-person learning	Binary i.e. 0 and 1

2.6 Summary

The literature above has shown that various studies have been completed through adoption of data-mining algorithms to evaluate learner performance and identify risk of incomplete grading. From the literature reviewed it is observed that though various studies have been conducted on the issues of missing marks, the data used (surveys and academic records) does not provide the necessary information to answer critical questions concerning student grading and outcome.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

In this section, I will capture the data gathering and analysis methodology that I will employ in this study. Areas to be discussed include study design, target group, research equipment, data collection techniques, data processing and analysis procedures. The research schedule and the budget are also proposed.

3.2 Research Design

The proposed design for this study is to use a quantitative approach focusing on exploratory research methodology. This serves well with the objectives, the conceptual framework and the research questions that are intended to be covered. Mixed-methods research design, consisting of both quantitative and qualitative data collection and analysis will also be essential.

The major purpose of exploratory research is to collect, analyze and identify patterns within the data, make predictions, test the relationships and extend the results to wider populations. The necessary information will be acquired from available datasets in kaggle and online questionnaires that will be given to the learner as a survey for additional information. The tools used to analyze this data will be python data analysis where the dataset will be imported analyzed for patterns and relationship. This will enable us gain critical understanding of how the variables have an effect on student incomplete grading.

The model approach of this study will be aligned with the mixed methods implementation approach as shown in the figure below. The purpose of the model is to provide the methods and a process of extracting knowledge from the datasets provided. Data collection from the learner will be the primary task. Data will then be divided into training and testing datasets, and the multivariate linear regression technique will be applied to the training dataset for prediction.

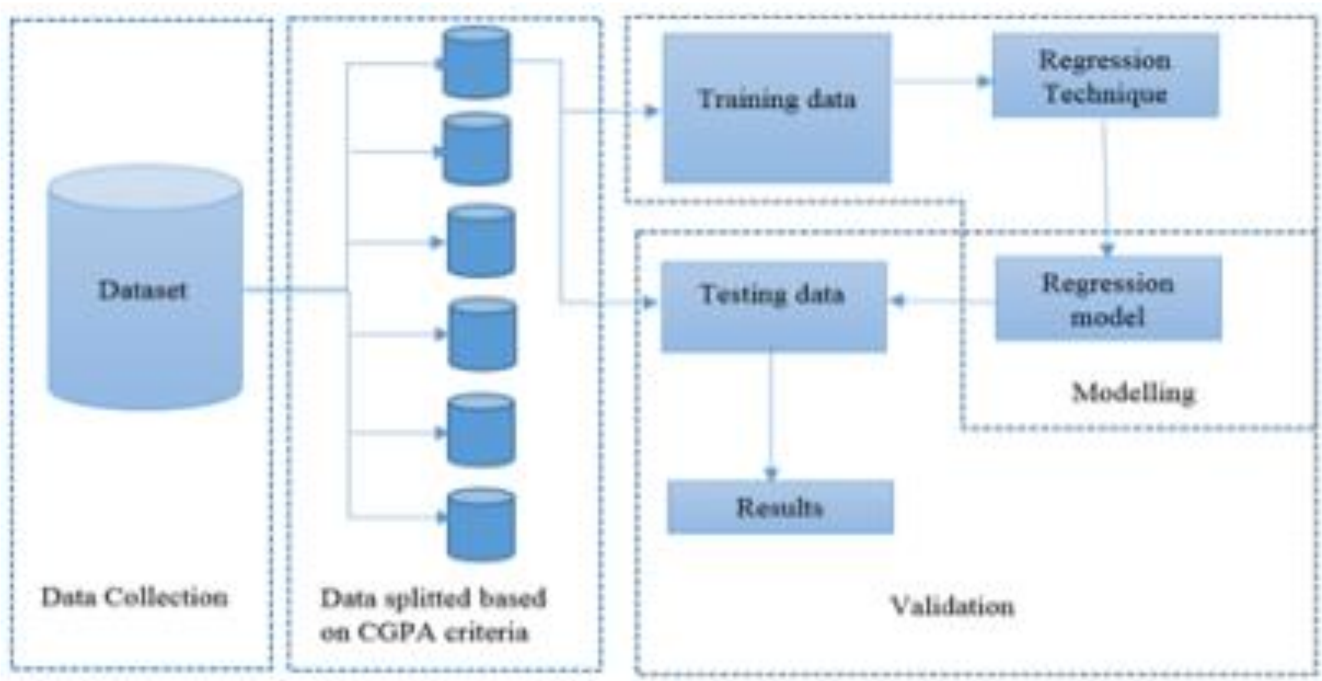

TABLE 3. 1: Flow of Predictive Model Implementation

Source: Rajalaxmi et al. (2019)

The next section describes the various stages in the design approach.

3.2.1 Data Collection

To collect data from the learners, a questionnaire will be created. Students enrolled in a variety of course disciplines will provide the input data on their academic experience with relation to missing marks.

3.2.2 Model validation

To see how well the model predicts and whether the test data is legitimate, more model validation should be performed.

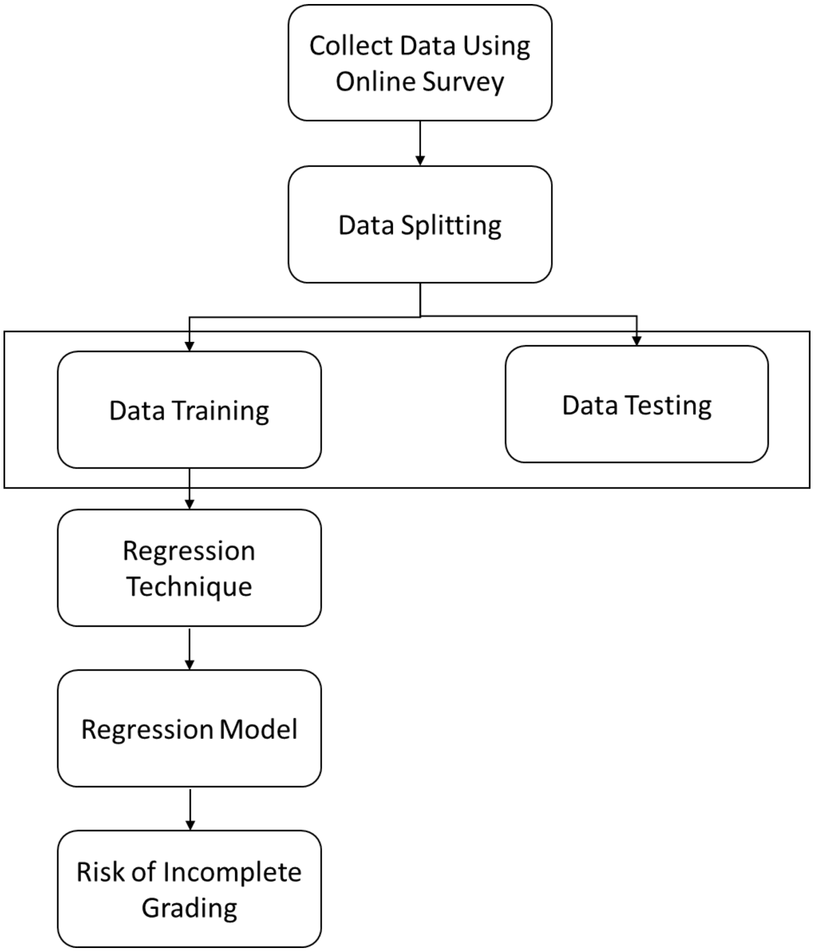

TABLE 3. 2: Flow Chart of the Proposed Method

3.3 Target population

Whenever a research is undertaken, one must identify the individuals on whom the results will be generalized (Jha, 2014). Subsequently, the population of the study comprises of the students with the desired shared qualities that inform the study. The target population that will be studied in the current research comprises of students from the Kenyan higher education sector. Some of the shared qualities in such a population is the fact that respondents are students studying in the selected private and public universities operating in Kenya. The population of students serves as the basis for the current study.

Consequently, the information gathered during this study from the sample will be generalized to the target population of students within the Kenya higher education sector. The target population of students within the Kenya higher education sector will form basis for external validity of the study.

The target population of this study are derived from undergraduate students within selected universities undertaking different courses between the periods of January 2022 to June 2023. The estimated total population will consist of 1,650 students with a sample size of 385 students will be used.

TABLE 3. 3: Sample Size

Category	**Population**	**Sample size**
Kca University	500	185
Mt. Kenya	150	50
Kenyatta University	500	50
JKUAT	500	100
Total	**1,650**	**385**

3.4 Sampling and Sampling Procedure

Once the target population is identified the next step is to select a pool of respondents that will participate in the study. It is impossible for the researcher to collect data from all the individuals within the banking sector due to the cost of such an event. Sampling is important since it will allow the researcher to focus on a particular subset of the target population (Daniel, 2012). The current study will use random sampling to determine the participants of the study.

This researcher was able to adopt simple random sampling method for the student-based sample in which a systematic random sample of students is to be drawn using an equal probability selection method. A simple random sample has been used to ensure that each sample of the population provided has an equal probability of being selected. This will guarantee the privacy of the student information utilized in this study.

3.4.1 Simple Random Sampling Formula

Calculation of sample size

Formula

Sample Size $n = t\hat{}2 * p(1-p)/m^2$

Where:

n = the sample size

t = the 95% confidence level which is equal to 1.96

p = the measure of prevalence

m = the 5% margin of error which is 0.05

Calculating the sample size portion to use

p	**(1 - p)**	**p (1 - p)**
1	0	0
0.7	0.3	0.21
0.6	0.4	0.24
0.5	0.5	0.25
0.4	0.6	0.24

From the above calculation we can conclude that the best sample portion is at 0.5 which gives a higher value of 0.25 for p (1-p).

n = (1.96) ^2 * (0.5) (0.5)/ (0.05) ^2

n = 384.16

n = **385** (Required sample size)

Based on the above formula, the ideal sample size will be 150 from a required sample of 385 of students who have enrolled to the LMS virtual campus.

3.4.2 Modelling the regression formula

From this formula…

$$\mathbf{y} \quad = \quad \boldsymbol{\beta_0 + \beta_1 x_1 \ (+\varepsilon)}$$

We have this formula…

Risk of Incomplete Grading | Student Characteristics | Student Assessment | Course Characteristics | Instructor Factors

$$\mathbf{y} \quad = \quad \boldsymbol{\beta_0} + \quad \boldsymbol{\beta_1 x_1} + \quad \boldsymbol{\beta_2 x_2} + \quad \boldsymbol{\beta_3 x_3} + \quad \boldsymbol{\beta_4 x_4 \ldots (+\varepsilon)}$$

Where,

y:	dependent variable
$\mathbf{x_1, x_2, x_3}$:	independent variables
$\boldsymbol{\beta_0}$:	y-intercept
$\boldsymbol{\beta_1, \beta_2, \beta_3}$:	parameters relating x's to y

3.5 Analysis

This study was based on descriptive analysis and predictive analysis to evaluate the effect of the multiple linear regression model.

3.5.1 Descriptive analysis

Based on the collected data from the conceptual framework variables, the researcher was able to calculate summary statistics for variables selected. This results include the mean, median, standard deviation, minimum, and maximum marks earned. The researcher was also able to examine the frequency and percentage of missing marks across different assessments, courses.

3.5.2 Predictive analysis

In this analysis, the study will use the multiple linear regression model to evaluate the effectiveness of the variables selected. To evaluate the validity and usefulness of the model R^2 will be used where R^2 = total variance / explained variance and this will help check how well the model fits the data.

To check whether the model (multilinear regression) is fitted efficiently, R^2 is performed (it is sometimes called adjusted R^2), defined as

$$R^2_c = R^2 - \frac{J(1-R^2)}{N-J-1}$$

Where:

J independent variables

N sample size

Akaike information criterion (AIC) method will also be used for evaluating how well the model fits the data it was generated from. According to AIC, the best model explaining as much variation as possible is one with a very small number of independent variables.

The AIC formula is as follows:

$$AIC = 2K - 2\ln(L)$$

where,

K is the total number of independent variables used (**Student Characteristics, Student Assessment, Course Characteristics, and Instructor Factors**)

L is log-likelihood estimate (the likelihood that the model could have produced the observed y-values).

The default K is equal to 2, if one independent variable is used in the model then the value of K is 3. The study uses 4 independent variables so our value of K will be 6.

3.5.3 Analysis of residuals

This study will check the difference between a variable's observed value and the variable's predicted value based on the multiple regression model. This will measure how far away a point is from the regression line. This will also help analyze the validity of the model.

Residual = Observed Value-Predicted Value

$$\boldsymbol{e = y - \hat{y}}$$

Where,

e the error term

y the observed value (independent variable)

ŷ is the predicted value (dependent variable)

3.6 Research Instrument

This study is based on data collected from components such as online survey questionnaires and compares the performance of students. The research instrument applied will be the use of EDM (Educational Data Mining) of relationship mining such as correlation and regression analysis. This will enable us identify patterns, correlation and association.

The study utilized a questionnaire with the view of obtaining relevant data consistent with the study objectives. Questionnaires serve as the best and appropriate tools for the collection of primary data while undertaking survey studies within a large population of respondents (Byrnes, 2009). The study used close ended questionnaires consisting of Likert scale in order to generate quantitative data.

To facilitate collection, analysis and preprocess of the data, python will be utilized as a general purpose programming language. I have selected python due to the fact that I have been able to familiarize with it during class sessions.

Pilot Testing

In order to determine the validity and reliability of the questionnaire, the study was taken through a comprehensive pilot test. The pilot test was undertaken on a few education students within Nairobi County. Khanchandani (2022) asserts that a pilot study entails the pre-testing of the research instrument; in this case a questionnaire, with the intention of heightening efficiency of the entire process. A sample size of 10-20% of respondents from the target population was sufficient for the pilot test. Subsequently, the current study used a sample of 8 respondents from the banking sector to run the pilot test. This pilot study was beneficial in determining how to administer the measuring instrument for conducting the actual study and will bring transparency to the research process. The study embarked on a new paradigm in reporting pilot studies, as the pilot study are usually under-discussed and less reported. The pilot study is the original work carried paved the way for the researchers, particularly to library and information professionals, before undertaking any survey study.

3.6.1 Validity Test

A research instrument should be able to accurately measure the intended objects and provide the correct information (Falk, 2022). Content validity of the research instrument was established through the use of Kaiser-Meyer-Olkin (KMO) and Bartlett's Test of Sphericity consultation with research experts. KMO test indicates the fraction of variance in the variables caused by the underlying factors. A value close to 1.0 is desirable. On the other hand, Bartlett's test of Sphericity helps determine whether factor analysis can be sufficient with the data. Table 3.1 shows that KMO value exceeds 0.5 implying that the sample is sufficient for factor analysis. Moreover, the Bartlett's Test of Sphericity indicates a value less than 0.05 of significance level. The results show that the research instrument used in data collection had content validity.

TABLE 3. 4: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.733
Bartlett's Test of Sphericity	Approx. Chi-Square 67.723
	df 6
	Sig. <.001

3.6.2 Reliability Test

The reliability test is integral in determining the internal consistency of the research instrument (Scott, 2019). Subsequently, the study used the Cronbach Alpha test to determine the internal consistency. It was imperative that a questionnaire produced consistent results once administered on different samples drawn from the same target population. A questionnaire that passes the reliability test should have a value greaterthan 0.7. On the other hand, a value below 0.7 indicates that the questionnaire needs to be improved.

TABLE 3. 5: Reliability Analysis

Variable	Test Items	Alpha Values
Instructor Factors	4	0.756
Student Factors	4	0.738
Course Factors	4	0.766

A comprehensive pilot study was undertaken with a view of determining the reliability of the questionnaire. Using the Cronbach Alpha metric, the internal consistency of the questionnaire was determined and it established that the four study variables questions were reliable.

According to table 5, Instructor factors had a reliability of 0.756; Student factors 0.738; Course factors had 0.766. Since the study variables had alpha values exceeding 0.7, the questionnaire was deemed as reliable.

TABLE 3. 6: **Cronbach Metric**

Reliability Statistics	
Cronbach's Alpha	N of Items
.857	44

3.7 Data Collection Procedure

Once the pilot test was complete, the required correction was undertaken, and a consent for data collection sought from the authorities. A formal letter of introduction from KCA University was obtained and served to introduce the researcher to the target population. Similarly, a letter of consent from the selected universities was sought to ensure that the researcher has authority to access the respondents in their places of work. The questionnaires were administered through the email in the different institutions in the month of July 2023. Following the administering of the questionnaires, I followed up on the progress of filling the data collection instrument on a weekly basis. Finally, in early July 2023 the questionnaires were collected from the same email after a period of one month had elapsed. The questionnaires collected were in a good state and clean hence ensuring that data recording and input was correct.

3.8 Data Processing and Analysis

The quantitative data was analyzed in order to come up with findings and conclusion. Once the questionnaires were returned, the researcher utilized the Statistical Package for Social Sciences (SPSS v27) to analyze the data. The data was analyzed using inferential statistics. The inferential statistics included correlation statistics and multiple regressions analysis. Multiple regressions is an example of ordinary least-squares model since it entails multiple explanatory variables. The multiple regressions function used in the study is illustrated below:

$Y= \beta_0 + \beta_1X_1+ \beta_2X_2+ \beta_3X_3+ \beta_4X_4 +\varepsilon$ Where:

Y	=	Risk of Incomplete Grading
β_0	=	Constant
X_1	=	Instructor Factors
X_2	=	Student Factors
X_3	=	Course Factors
X4	=	Student Assessment
ε	=	ErrorTerm

3.9 Ethical Consideration

The research process was guided by ethical guidelines including informed consent, and respect for respondents' confidentiality. Moreover, the research respected the privacy of the respondents and the organizations they represented. In this regard, the data collection process involved unmarked questionnaires that did not indicate the respondent's personal details.

Following the collection of data from learners, the study ensured the security of student privacy and confidentiality when addressing the issue of incomplete grading and missing marks in the education sector. Student cademic records and personal information was treated with the highest care and in accordance with applicable data protection laws and regulations.

CHAPTER FOUR

DATA ANALYSIS, FINDINGS AND DISCUSSIONS

4.1 Introduction

The following chapter of the research commences with an illustrative description of the findings drawn from collection of data. The data was collected from 367 respondents who represent 92.3% of the sample population. Subsequently, within this chapter, data is presented through the form of distribution tables that aid in the description of findings coupled with the relevant explanation. The discussion of findings was meant to attain the specific objectives set for the study.

This chapter presents the results and analysis of the study. The data was acquired from 367 higher education students in Kenya. This sample was guided by (Bakri, 2023) which states that quantitative studies require testing of at least 500 sample respondents. (Lund, 2023) emphasize that the recommended sample sizes should be closer to 100 for descriptive research design.

From the data analysis, the students were drawn from seven different courses from four universities in Kenya. Three of the universities were public universities and one private mainly KCA University. There were 367 respondents where students pursuing Information Technology had 55 respondents, 28 respondents were from Commerce, 39 respondents were from Business Management, 41 were from Economics, 179 from Education, 17 from Data Science and 8 from Medicine. The male respondents were 207 compared to female respondents who were 160.

Questionnaires designed using google forms were administered to the respondents via what’s app and email technologies. In the design of the questionnaire we preferred to use Likert scale values of 1-5 than the percentages due to the fact that the values can support assessment of opinions, behaviors and attitudes (Karakose, 2023) in a more specific manner.

4.1.1 Response Rate for the Study

TABLE 4. 1: Response Rate

	Questionnaires Administered	Questionnaires filled & Returned	Percentage
Respondents	367	360	98.1 %

The researcher sought to gather data from students within the private and public universities and more specifically within KCA University. Subsequently, a response rate of 92.3% was achieved and this is satisfactory for generalization of the findings.

4.1.2 Demographic Information

TABLE 4. 2: Gender Distribution

	Frequency	Percentage (%)
Male	207	56.4
Female	160	43.6
Total	**367**	**100.0**

The findings illustrated on table 5 indicate that a majority of the respondents were male at 56.4%. on the other hand, 43.6% of the respondents were female. Subsequently, a majority of students in the university response are male.

4.1.3 Length of study

	Frequency	Percent	Valid Percent	Cumulative Percent
1.00 Less than one month	19	5.2	5.2	5.2
2.00 More than a month less than 6 months	16	4.4	4.4	9.5
3.00 More than one year	55	15.0	15.0	24.5
4.00 Less than 3 years	14	3.8	3.8	28.3
5.00 More than 3 years	263	71.7	71.7	100.0
Total	367	100.0	100.0	

TABLE 4. 3: Length of Study

According to table 6 a majority of the respondents have studied in the universities for more than three years (71.7%). The second most populous group of respondents are those students who have studied in the university for more than one year (15%). Similarly, respondents who have studied for less than one month are 5.2% of the entire population, respondents who have studied for more than a month or less are 4.4%, and the respondent who have been studying for less than three years are 3.8%.

4.1.4 Age category

	Frequency	Percentage
Below 20 Years	45	12.3
20 - 30 Years	231	62.9
31-40 Years	74	20.2
Above 41 Years	17	4.6
Total	**367**	**100.0**

TABLE 4. 4: Spread of Respondents by Age

According to table 7, a majority of the respondents (62.9%) are between the age of 20 and 30 years old. Similarly, 20.2% of the respondents are between 31 and 40 years old; those below 20 years make up 12.3%of the population. Finally, the minority population is that of respondents above 41 years of age at 4.6%.

4.1.5 Level of education

Course	**Frequency**	**Percent**
Postgraduate	64	17.4
Degree	76	20.7
Certificate	93	25.3
Diploma	134	36.5
Total	**367**	**100.0**

TABLE 4. 5: Level of Education

Table 8 shows that a majority of the respondents in the university have a first diploma qualification at 36.5% of the population. The second most populous category is college certificate with at least 25.3% of the respondents. The third most populous course category is college degree with having respondents of about 20.7%. Finally, 17.4% of the respondents have a postgraduate.

4.2 Research findings

4.2.1 Objective One Results

The first objective of the research study was to investigate the prevalence of missing marks and incomplete grading and its effect on learners in higher education as well as examine the factors that lead to incomplete grading. Subsequently, the research established from the respondents through the questionnaire that the following are the main factors:

i. Attendance record

This challenge was established from part B in the questionnaire. From table 4.7, we can see 151 respondents and 41.1% of the respondents strongly agree with large extent to the fact that class attendance is essential in managing missing marks. Students class attendance is important when analyzing missing marks issue.

I believe i have a good class attendance record					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	13	3.5	3.6	3.6
	2.00 Small extent	10	2.7	2.7	6.3
	3.00 Moderate extent	64	17.4	17.5	23.8
	4.00 Large extent	151	41.1	41.3	65.0
	5.00 Very large extent	128	34.9	35.0	100.0
	Total	366	99.7	100.0	
Missing	System	1	.3		
Total		367	100.0		

TABLE 4. 6: Class Attendance

ii. Communication to the instructor on missed assessment

From table 10, 167 respondents strongly agree with large extent and 79 respondents moderately agree with the fact that they have to communicate to the instructor on missed assessment is essential and failure to do so can lead to incomplete grading. This is a huge number compared to the respondents who agreed to have the knowledge. As a result, this challenge was noted. This data was sought from part E of the questionnaire.

I often communicate in case I miss an assessment					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	14	3.8	3.8	3.8
	2.00 Small extent	34	9.3	9.3	13.1
	3.00 Moderate extent	79	21.5	21.6	34.7
	4.00 Large extent	167	45.5	45.6	80.3
	5.00 Very large extent	72	19.6	19.7	100.0
	Total	366	99.7	100.0	
Missing	System	1	.3		
Total		367	100.0		

TABLE 4. 7: Assessment communication

iii. Group Work Assessment

From table 11, 126 of the respondents strongly agree with very large extent and 121 respondents largely agree with the fact that they prefer group work assessments given by the instructor instead of individual class assessments. We can conclude that in groups there is a higher probability of the student not having an incomplete grade as there is assistance granted by the group members as a whole. As a result, this challenge was noted. This data was sought from part E of the questionnaire.

I prefer group work assessment					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	16	4.4	4.4	4.4
	2.00 Small extent	28	7.6	7.6	12.0
	3.00 Moderate extent	76	20.7	20.7	32.7
	4.00 Large extent	121	33.0	33.0	65.7
	5.00 Very large extent	126	34.3	34.3	100.0
	Total	367	100.0	100.0	

TABLE 4. 8: Group work assessment

iv. Timely assessment feedback

From table 12, 121 of the respondents strongly agree with very large extent and 106 respondents largely agree with the fact that once the assessment has been given by the instructor, it is essential that feedback is provided in good time. This allows the student to plan well and shift focus on other assessments. We can conclude that timely feedback is essential once assessment has been given. As a result, this challenge was noted. This data was sought from part E of the questionnaire.

I am confident assignments are checked and feedback is given within an acceptable period of time					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	11	3.0	3.0	3.0
	2.00 Small extent	49	13.4	13.4	16.3
	3.00 Moderate extent	80	21.8	21.8	38.1
	4.00 Large extent	121	33.0	33.0	71.1
	5.00 Very large extent	106	28.9	28.9	100.0
	Total	367	100.0	100.0	

TABLE 4. 9: Timely Assessment feedback

v. Virtual Campus compliance

From table 13, 132 of the respondents strongly agree with in large extent and 118 respondents in very largely agree with the fact that they prefer assessments issued through virtual campus and not manual assessment. This can improve on accountability of knowing the assessment attempted and that is not attempted. This provides an opportunity for the student to plan well and communicate on missed assessments. We can conclude that timely feedback is essential once assessment has been given. As a result, this challenge needs to be addressed appropriately so as to enhance reduce cases of missed assessments. This data was sought from part E of the questionnaire.

The instructor often issues assignments on virtual campus					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	11	3.0	3.0	3.0
	2.00 Small extent	39	10.6	10.6	13.6
	3.00 Moderate extent	67	18.3	18.3	31.9
	4.00 Large extent	132	36.0	36.0	67.8
	5.00 Very large extent	118	32.2	32.2	100.0
	Total	367	100.0	100.0	

TABLE 4. 10: Virtual Campus compliance

vi. Course Materials

From table 14, 99 of the respondents strongly agree with in large extent and 51 respondents in very largely agree with the fact that they prefer course materials be provided by the instructor through virtual campus. This can improve on accountability of knowing the assessment attempted and that is not attempted. This provides an opportunity for the student to plan well and communicate on missed assessments. We can conclude that timely feedback is essential once assessment has been given. As a result, this challenge needs to be addressed appropriately so as to enhance reduce cases of missed assessments. This data was sought from part E of the questionnaire.

I am aware that this course has adequate resources to support learning provided by the lecturer					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	55	15.0	15.0	15.0
	2.00 Small extent	77	21.0	21.0	36.0
	3.00 Moderate extent	85	23.2	23.2	59.1
	4.00 Large extent	99	27.0	27.0	86.1
	5.00 Very large extent	51	13.9	13.9	100.0
	Total	367	100.0	100.0	

TABLE 4. 11: Course materials

vii. Course outline

From table 14, 111 of the respondents strongly agree in very large extent and 106 respondents in largely agree with the fact that they prefer course outline be provided before the prior to classes by the instructor through virtual campus. This can improve on accountability on knowing when the assessment will be issued at a given week. This provides an opportunity for the student to plan well and prepare for the assessments. We can conclude that the course outline is essential to managing missing marks. As a result, this challenge needs to be addressed appropriately so as to enhance reduce cases of missed assessments. This data was sought from part E of the questionnaire.

I believe this course outline is issued					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	16	4.4	4.4	4.4
	2.00 Small extent	40	10.9	10.9	15.3
	3.00 Moderate extent	94	25.6	25.6	40.9
	4.00 Large extent	106	28.9	28.9	69.8
	5.00 Very large extent	111	30.2	30.2	100.0
	Total	367	100.0	100.0	

TABLE 4. 12: Course outline

Viii. Punctuality

From table 16, 126 of the respondents strongly agree in large extent and 83 respondents in very largely agree with the fact that instructors need to be punctual enough when attending class sessions. This can improve on accountability on knowing when the assessment will be issued at a given week. This provides an opportunity for the student to plan well and prepare for the assessments. We can conclude that the course outline is essential to managing missing marks. As a result, this challenge needs to be addressed appropriately so as to enhance reduce cases of missed assessments. This data was sought from part E of the questionnaire.

I believe the lecturer is punctual					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	12	3.3	3.3	3.3
	2.00 Small extent	45	12.3	12.3	15.5
	3.00 Moderate extent	101	27.5	27.5	43.1
	4.00 Large extent	126	34.3	34.3	77.4
	5.00 Very large extent	83	22.6	22.6	100.0
	Total	367	100.0	100.0	

TABLE 4. 13: Instructor Punctuality

Ix. Make up issued by the instructor

From table 17, 140 of the respondents strongly agree in large extent and 102 respondents in very largely agree with the fact that instructors need to issue make up assessments incase the student misses an assessment. This can improve on accountability on knowing when the assessment will be issued at a given week. This provides an opportunity for the student to plan well and prepare for the make-up assessments and reduce any chances of having an incomplete grade. We can conclude that the course outline is essential to managing missing marks. As a result, this challenge needs to be addressed appropriately so as to enhance reduce cases of missed assessments. This data was sought from part E of the questionnaire.

Make up is issued in case i miss an assessment					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	19	5.2	5.2	5.2
	2.00 Small extent	28	7.6	7.6	12.8
	3.00 Moderate extent	78	21.3	21.3	34.1
	4.00 Large extent	140	38.1	38.1	72.2
	5.00 Very large extent	102	27.8	27.8	100.0
	Total	367	100.0	100.0	

TABLE 4. 14: Issuing of Make up

X. Course planning

From table 21, 111 of the respondents strongly agree in large extent and 83 respondents in very largely agree with the fact that courses need to be well planned enough prior to the class sessions. This can improve on accountability on knowing when the assessment will be issued at a given week. This provides an opportunity for the student to plan well and prepare for the assessments. We can conclude that the course outline is essential to managing missing marks. As a result, this challenge needs to be addressed appropriately so as to enhance reduce cases of missed assessments. This data was sought from part E of the questionnaire.

TABLE 4. 15: Course Planning

I believe this course is well prepared					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not at all	16	4.4	4.4	4.4
	2.00 Small extent	40	10.9	10.9	15.3

	3.00 Moderate extent	94	25.6	25.6	40.9
	4.00 Large extent	106	28.9	28.9	69.8
	5.00 Very large extent	111	30.2	30.2	100.0
	Total	367	100.0	100.0	

4.2.2 Objective Two Results

The second objective of the research study was to examine the factors that promote the chances that lead to incomplete grading in higher education. The results for the second objective are illustrated in table 4.16.

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.565	.125		4.503	.000
	Studentfactors	.092	.030	.075	3.025	.003
	courseeffects	-.060	.024	-.067	-2.503	.013
	Assessmentfactor	.861	.024	.891	36.565	.000
	instructorfactor	-.071	.023	-.085	-3.111	.002
a. Dependent Variable: Incompletegrading						

TABLE 4. 16: Results for objective two

Subsequently, following the analysis, the researcher established that Instructor Factor, Student Factor, Assessments factors and Course effects are statistically significant in the attainment of incomplete grading when using regression technique in the education sector. Moreover, the table indicates that as Student factors index changes by a value of 1, it leads to a change of 0.092 change in sustainable prediction of incomplete grading of students. Similarly, if the Course effects index changes by value of 1, then it leads to a -0.060 change in the poor sustainable prediction in incomplete grading of students. On the other hand, if Assessment factors index changes by a value of 1, it leads to a change of 0.861 change in sustainable prediction of incomplete grading of students. On the other hand, instructor factor has a negative value of -0.071 implying a change in the existing instructor culture will lead to a poorer corporate management of incomplete grading within the higher education sector.

4.2.3 Objective Two Results

The objective was to design and develop a Model that will predict the risk of incomplete grading based on student assessment in higher education.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$ Where:

Y	=	Incomplete Grading
β_0	=	Constant
X_1	=	Student Factors
X_2	=	Course Factors
X_3	=	Student Assessment Factors
X_4	=	Instructor Factors
ε	=	Error Term
$\beta_1, \beta_2, \beta_3, \beta_4$	=	Regressions Coefficients

Therefore, Incomplete Grading = 0.565+0.092 Student factors-0.060 Course effects+0.861 Assessment factors-0.071 Instructor factors.

The developed model shows that student factors, course effects, Assessment factors and instructor factors have a significant association with managing incomplete grading within the Kenya Higher Education sector. Subsequently, changes in the value of the variables affect management of incomplete grading when integrating regression model in the education sector.

4.2.4 Objective Three Results

The third objective of the research study was to test and validate the model. Subsequently, Model Summary and ANOVA were used. The researcher developed an OLS model using multiple regressions to test the impact of the predictor variables on the dependent variable. Subsequently, SPSS v27 was used to input, code, and compute multiple regressions statistics. The model summary is illustrated in table 4.17.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.907[a]	.823	.821	.28456
a. Predictors: (Constant), instructorfactor, Assessmentfactor, Studentfactors, courseeffects				

TABLE 4. 17: Model Summary

a. Predictors: (Constant), courseeffects, instructorfactor, Studentfactors

As illustrated in table 23, the correlation existing between the independent variables (Course effects, Instructor factors, Student factors and Assessment factors) and the dependent variable of predicting incomplete grading of students in higher education (R=0.907). On the other hand, the Adjusted R- Square which is the coefficient of determination (r^2=0.82) indicates an 82% change in predicting incomplete grade of students in higher education under linear regression attributable to the three predictor variables. In this regard, there are a myriad of other factors affecting incomplete grading and this needs to be explored.

ANOVAa						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	136.434	4	34.109	421.239	.000[b]
	Residual	29.312	362	.081		
	Total	165.746	366			
a. Dependent Variable: Incompletegrading						
b. Predictors: (Constant), instructorfactor, Assessmentfactor, Studentfactors, courseeffects						

TABLE 4. 18: ANOVA

On the other hand, table 24 depicts the Analysis of variance results which has been used in the study for the purpose of determining whether there are any statistically significant differences between the mean/average of the four independent variables. The main limitation of the ANOVA test statistic is the fact that it cannot describe which specific groups in the data set that are statistically different from each other (Hirotsu, 2017). In the current ANOVA analysis, the generated significance value is 0.001 which alsoimplies a p-value of .001 that is below 0.05. Consequently, there is a statistically significant difference between the means of the independent variables Student factors, instructor factors, Course effects and Assessment factors. The limitation of ANOVA test statistic prevents us from determining the specific variables that differed.

4.3 Discussion of Results

The research study involved 367 respondents from the selected Kenyan private and public universities sector who have experienced incomplete grading in their various institutions. At 5% level of significance, it was established that the three independent variables chosen – Student factors, Instructor factors and Course factors - had a significant effect on predicting incomplete grading. Subsequently, the data collection process reviewed the variables of the study and established that the management of student assessments within the organization is imperative.

The general objective of the research was to establish a model for predicting incomplete grading in the Kenyan private and public universities. Subsequently, three aspects were reviewed to determine their impact on incomplete grading.

Subsequently, the data collection process reviewed the variables of the study and established that the management of missing marks used in the universities is imperative. Such findings correlate with those of Chao et al. (2020) who posit that progressive organizations have integrated various systems in the workplace and management of the incomplete grading of students. Moreover, Jiunn-Woei (2020) affirm the importance of managing missing marks management through computing technology and the use of proper systems.

Similarly, management policies were identified as a significant aspect of managing incomplete grading in the study. Such policies are meant to guide the leaner on how to mitigate against risks that can lead to having an incomplete grade. The study findings indicate that management policies are integral in the management of incomplete grading such as aligning the policies towards the instructor factors, student factors and course factors. The findings are consistent with those by Chen et al. (2022) who stated that novel technology platforms are linked with student and assessment risks. Subsequently, the establishment of missing marks policies is critical in the higher education sector. On the other hand, the study established that changes in the instructor factors in the organization can lead to diminishing risks of incomplete grading (Grant, 2022). The current findings indicate that instructor factors relating to the instructor can contribute towards decline in risk in incomplete grading regardless of the age demographic of users. The majority of the respondents (62.9%) were between 20 and 30 years of age which is a youthful population.

Conversely, the research established that student’s assessment feedback was significant in developing a sustainable model to predict the risk of incomplete grading. The results are consistent with the (Bostedt, 2022) social cognitive theory to identify whether student motivation and lecturer teaching contribute to how a student performs. The study reveals that student motivation to study is based on home situation (external factor) and student driving force (internal factor). The theory provides that once technology learners perceive a threat, they will become motivated to overcome and avoid the threat. Subsequently, leaners who perceive incomplete grading threats are more likely to accept education and training for better planning hence improving the management of incomplete grades. Learner training and education heightens the knowledge held by the respondents when perceiving incomplete grades.

4.4 Summary

Following the comprehensive data analysis, the study established that the three independent variables of student factors, instructor factors, assessment factors and course factors have a significant effect on the predicting the risk of incomplete grading. Subsequently, integration of having proper systems and policies in the education sector requires to be guided the following regression model: Incomplete Grading = 0.975+0.267 Student factors+0.556 Course effects-0.073 Instructor factors.

4.5 Results for research question one

The specific objective one of this study was to find attributes that can be used to predict the risk of incomplete grading of students in higher education. From the literature review three attributes were identified including.

4.5.1 Instructor characteristics

This attribute considers how the students grading is affected with reference to instructor indicators such as lecture plans, lecturer punctuality, adequate knowledge on course, teaching style, timely feedback on assessments, teaching approach, preparedness of the lecturer during classes, timely feedback by the instructor to student’s work and whether the instructor guided the students to complete their course on time allocated.

Instructors are responsible for grading assignments and exams. If an instructor does not grade assignments in a timely manner, it can result in incomplete grading. Students may not receive feedback or grades promptly, leading to uncertainties about their performance. Grading practices should be consistent and aligned with the established criteria. Inconsistent grading can lead to disputes and requests for reevaluation, contributing to incomplete grading.

Instructors should maintain clear and open communication with students. If students have questions about their grades or grading criteria and the instructor is not responsive, it can create grading issues. Instructors are often responsible for entering grades into digital systems. Errors in data entry can lead to missing or incorrect grades, resulting in incomplete grading. Instructors play a critical role in setting clear expectations for assignments and exams. Unclear instructions can lead to misunderstandings, disagreements, and rework requests, potentially causing grading delays and incompleteness.

The complexity and design of assignments can affect grading. If assignments are too complex or time-consuming, instructors may require more time to grade them, potentially leading to incomplete grading. Instructors should be accessible to students for questions and concerns related to grading. If students cannot reach their instructors, unresolved issues can contribute to incomplete grading. Constructive and meaningful feedback is an integral part of grading. If instructors do not provide detailed feedback, it may result in grading disputes and requests for clarification, leading to incomplete grading.

Instructors can support students by offering guidance on how to improve their work and by providing opportunities for revisions. If this support is lacking, students may be less likely to complete assignments adequately. Instructors should follow clear and well-documented grading policies and procedures. When grading practices are inconsistent or ambiguous, it can lead to grading disputes and incompleteness.

4.5.2 Student characteristics

This attribute considered how the students grading is affected with reference to students' capability to help improve their abilities in the course activities. The indicators include whether the student attempted assessment issued by the instructor, has a good attendance record, participates in group work, communicates on missed assessments, has fee issues during study, attempts makeup assessment when issued by the instructor.

Students may submit assignments, projects, or exams late or not at all. Late submissions or non-submissions can lead to grading bottlenecks and incomplete grading.

Students may request reevaluation of their grades if they believe that they were graded unfairly. This can lead to additional grading work for instructors and potential delays in finalizing grades. If students do not promptly communicate with instructors or administrative staff regarding grading concerns or discrepancies, it can lead to grading issues not being addressed in a timely manner. Students may submit incomplete or partially completed assignments or exams, requiring instructors to make additional assessments or request revisions.

Students are often responsible for entering or verifying their personal information, including their student ID numbers, in digital systems. Incorrect data entry by students can result in data discrepancies and potential grading issues. If students dispute their grades or appeal grading decisions, it can lead to a more prolonged grading process and potential incomplete grading until disputes are resolved.

Misunderstandings about assignment requirements or grading criteria can lead to student dissatisfaction with their grades. In such cases, students may request reevaluation or clarification, which can slow down the grading process. The quality and completeness of student submissions can influence the time required for grading. Poorly organized or unclear work may take longer to assess, potentially leading to grading delays.

4.5.3 Course characteristics

This attribute considered how the students grading is affected with reference to course indicators such as course difficulty, availability of the course outline. Students taking a heavy course load often have a greater number of assignments, exams, and projects to complete. As a result, they may experience difficulties in managing their workload, potentially leading to late submissions or incomplete assignments. Educators, too, may struggle to handle grading for numerous assignments, potentially causing delays.

Students taking a full course load often have limited time to dedicate to individual courses and assignments. This can lead to rushed or less thorough work, which may affect the quality of their submissions and, consequently, the grading process. In larger classes or with a high number of students, communication can become more challenging. Students may find it difficult to reach out to educators with grading concerns, and educators may struggle to provide timely feedback or address questions.

From the educators' perspective, having a large number of students in a course can increase the grading workload. This can lead to grading delays and potentially incomplete grading if educators are unable to keep up with grading for all students. In courses with many assignments, exams, and projects, students may face multiple deadlines in a short time frame.

Meeting these deadlines can be challenging, and late submissions are more likely to occur, contributing to incomplete grading.

In courses with a high student-to-educator ratio, the quality of feedback on assignments and exams may be compromised. Educators may have less time to provide detailed and constructive feedback, impacting students' learning and the grading process. In courses with a large number of students, the digital infrastructure and learning management systems may experience technical challenges. This can result in difficulties in submitting assignments, accessing grades, or processing grading data. With a heavy course load, students must be highly accountable for their academic responsibilities. If they fail to keep track of assignments, deadlines, and grading, they may be more likely to encounter incomplete grading issues.

4.5.4 Student Assessment characteristics

Student assessment practices are closely intertwined with the occurrence of incomplete grading in higher education. The way assessments are designed, administered, and managed can directly impact the likelihood of incomplete grading. Clear Assessment Guidelines: Unclear or ambiguous assessment guidelines can lead to disagreements or confusion between students and educators. If students seek clarification or revisions, this can delay the grading process.

Complex assessment methods, such as long essay assignments, detailed projects, or intricate practical exams, can increase the chances of incomplete grading. Educators may require more time to assess these assignments, and any delays can lead to incomplete grading. A high volume of assessments, especially in large classes or during peak assessment periods, can overwhelm educators. This may lead to grading delays or errors, contributing to incomplete grading.

Students' timely submission of assessments is crucial. Late submissions may create grading bottlenecks and result in incomplete grading. Educators may prioritize grading on a first-come, first-served basis. Unclear or ambiguous assessment guidelines can lead to disagreements or confusion between students and educators. If students seek clarification or revisions, this can delay the grading process. Incomplete or delayed feedback on assessments can impact students' motivation and engagement. When students do not receive timely feedback, they may lose interest in their studies or fail to identify areas for improvement.

The use of digital platforms and learning management systems for assessments can introduce technological challenges. Technical issues, such as system crashes, data loss, or access problems, can result in incomplete grading. The workload of educators is a critical factor. Overworked educators may struggle to manage their grading responsibilities, leading to delays or errors in grading, which can contribute to incomplete grading. In institutions where educators are responsible for manual data entry of grades into digital systems, data entry errors can result in incomplete grading. Forgetting to enter grades or entering incorrect data can lead to missing grades. Effective communication with students is essential. If students do not understand the grading process, expectations, or deadlines, they may be more likely to experience incomplete grading issues.

Grading policies at the institutional level play a significant role. The absence of clear grading policies or inconsistent practices can result in incomplete grading. Institutions that lack standardized grading procedures are more susceptible to this issue. Students' responsibility for ensuring that their assignments are submitted and graded accurately is also a factor. If students do not keep records of their submitted work or communicate grading concerns to educators, incomplete grading issues may persist.

The availability of student support services, such as academic advising, can impact students' ability to address grading issues and advocate for accurate and timely grading. That is, measures of social and academic engagement, such as low grades, misbehavior, and high absenteeism, predicted both whether students changed schools or dropped out.

4.5.5 Organization characteristics

This attribute considered how the students grading is affected with reference to organization indicators such as addressing issues relating to missing marks, whether there are proper systems in place to manage issues with incomplete grading. The organization's grading policies, including guidelines on how grading should be conducted and deadlines for grade submission, can directly influence grading practices and the potential for incomplete grading.

Organizations can implement quality control mechanisms to review grading practices and ensure consistency and accuracy. This can help identify and rectify incomplete grading issues. The technological infrastructure provided by the organization, including learning management systems and grading software, can impact the grading process. Technological challenges or inefficiencies may contribute to incomplete grading. The availability of student services, such as academic advising and support for resolving grading concerns, can affect students' ability to address incomplete grading issues.

4.6 Correlation Results

In order to determine the relationship between the variables under study, the study used Karl Pearson's product moment correlation analysis. The findings were as shown in the table 4 below. The correlation is perceived to be significant when the probability value is below 0.05(p-value less than 0.05). correlations value (r) close to zero means a weak relationship and r close to one mean a very strongly correlations existing Table 4 below represents the results of the correlations analysis.

TABLE 4. 19: Correlation Matrix for Objective One

Correlations						
		Incomplete grading	Student factors	instruct orfactor	course effects	Assessmen tfactor
Incompletegradin g	Pearson Correlation	**1**	.406**	.040	.051	.897**
	Sig. (2-tailed)		.000	.443	.330	.000
	N	367	367	367	367	367
Studentfactors	Pearson Correlation	.406**	**1**	.272**	.212**	.413**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	367	367	367	367	367
instructorsfactors	Pearson Correlation	.040	.272**	**1**	.563**	.160**
	Sig. (2-tailed)	.443	.000		.000	.002
	N	367	367	367	367	367
courseeffects	Pearson Correlation	.051	.212**	.563**	**1**	.168**
	Sig. (2-tailed)	.330	.000	.000		.001
	N	367	367	367	367	367
Assessmentfactor	Pearson Correlation	.897**	.413**	.160**	.168**	**1**
	Sig. (2-tailed)	.000	.000	.002	.001	
	N	367	367	367	367	367
**. Correlation is significant at the 0.01 level (2-tailed).						

The results indicate that Incomplete grading and the student factor are positively and significant related (r=0.406, p=0.000) the relationship is fully supported by finding by Abou and Mehdi (2023) study was on effect of student performance on incomplete grading. The study focused on collecting qualitative information from the student's GPA. The information collected was analyzed qualitatively. The study found out that student attendance record, previous grades and gender is of critical importance for determining the risk of incomplete grading.

The table further indicated that Incomplete grading and Instructor factor are positively and significantly related (r=0.040, p=0.443), the relationship is also supported by (Wright, 2022) conducted a study on student incomplete information on academic performance. The study determined that students may only learn about course difficulty after performing poorly on an exam later in the semester. This study aimed to evaluate the effect on incomplete grading of course variables using regression model for a survey of fifty (50) participants. We found a strong association between course effects and student incomplete grading. It was also established that Incomplete grading and Course effects were positively and significant related (r=0.051, p=0.330).

It was also established that Incomplete grading and Assessment factors were positively and significant related (r=0.897, p=0.000).

4.6.1 Results for research question one

The specific objective one of this study was to investigate the prevalence of missing marks and incomplete grading and its effect on learners in higher education as well as examine the factors that lead to incomplete grading. From the literature review missing marks has had a bigger effect on student's progress as 72% of students have considered dropping out of school as they don't value repeating the unit.

The study confirms that academic performance (assessments), attendance record and fee issues factors as the highest predictors of student incomplete grading. Academic reasons such as student motivation were significant, as vocational motivation and goals established by student-instructors early in their education help prevent incomplete grading. Contrary to expectations, student-teachers' cultural values, parents' level of education, and cost of financing education had no significant impact on incomplete grading. This is most likely due to the Government's financial support for most students the widespread belief that higher education can improve one's social and economic status. The findings indicate that early identification and dropout prevention efforts should integrate various support services to foster a healthy learning and retention environment.

The specific study was to identify the factors that can affect incomplete grading in higher education. From the literature review missing marks can be due to a variety of factors. This research used predictive analytics by analyzing various independent variables that affect incomplete grading in higher education.

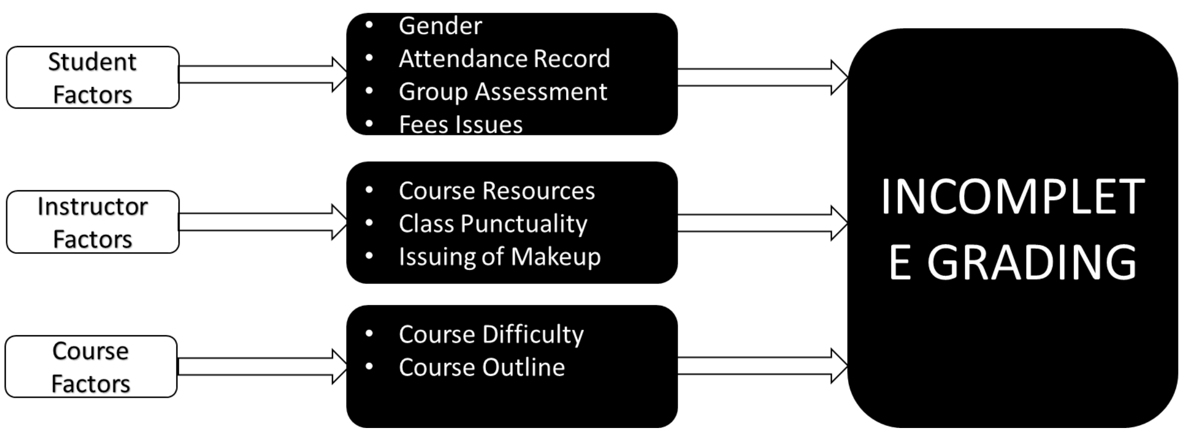

FIGURE 4. 1: Variables Used to Analyze the Effect of Incomplete Grading

From the figure (5) above, student factors vary from gender, attendance record of the student, participation in group assessments and having fee issues. This factors play a major role in student having an incomplete grade. Based on the insights from the analysis done using python, we identified that there was a balanced response from both male and female as we recorded a male response of 207 (56.40%) and a female response of 160 (43.60%) as seen in the figure (6) below.

Student attendance has played a major factor in student performance. Alsariera et al. (2022) in a review conducted, it was discovered that overall grade averages and internal assessments (quizzes and attendance) are commonly used in predicting student results which can lead to incomplete grading. The figure below shows an insight of student attendance for the analyzed data.

As we have observed from the figure (7) below, 41.14% large extent of students have an attendance issue which results to missing out on assessments issued by the instructor. This can lead to the student having an incomplete grade. Most students in large extent of 34.33% prefer group assessments when being graded with a small number preferring individual assessment with 4.36%. This reduces the chance of having a missing grade. Group assessments allows the student to participate in a collaborative way with his or her colleagues as it shifts all the burden of doing an assessment alone. Most students feel comfortable with group assessments as there is assistance distribution among the peers.

From this research, we have identified fee issues as another critical factor that can lead to student having an incomplete grade. Students tend to miss a lot of class sessions due to circumstances related to fee as they have to stop the session and resume or repeat the course due to not completing all coursework issued by the instructor. About 38.69% believe fee is their concern to having an incomplete grade with 1.91% experiencing low issues with fee payment.

4.6.2 Results for research question three

The specific objective two of this study was to identify the appropriate model we can use to measure the risk of incomplete grading in higher education. From this literature, we analyzed the data using different models which included Linear Regression, Lasso, K-Neighbors Regressor, Decision Tree, Random Forest Regressor, Gradient Boosting, XGBRegressor, CatBoosting Regressor, AdaBoost Regressor. In order to determine the appropriate model to use, we compared the model's performance training set and performance test set.

Model Training and Testing

Model Performance for Training set

TABLE 4. 20: Model Performance Training

Model	RMSE	MSE	MAE	R2 SCORE
Linear Regression	5.3231	28.3349	4.2667	0.8743
Lasso	1.1280	1.2724	0.9757	0.000
K-Neighbors Regressor	0.7221	0.5214	0.5700	0.5902
Decision Tree	0.0000	0.0000	0.0000	1.0000
Random Forest Regressor	0.2763	0.0763	0.2124	0.9400
Gradient Boosting	0.3275	0.1072	0.2629	0.9157
AdaBoost Regressor	0.7088	0.5024	0.6158	0.6051
XGBRegressor	0.0013	0.0000	0.0008	1.0000
CatBoosting Regressor	0.0398	0.0016	0.0303	0.9988

Model Performance for Test set

Model	RMSE	MSE	MAE	R2 SCORE
Linear Regression	5.3940	5.3940	4.2148	0.8804
Lasso	1.0165	1.0332	0.8450	-0.0001
K-Neighbors Regressor	0.7725	0.5968	0.6054	0.4223
Decision Tree	1.1390	1.2973	0.6757	-0.2558
Random Forest Regressor	0.7292	0.5317	0.5011	0.4853
Gradient Boosting	0.7367	0.5427	0.5102	0.4747
AdaBoost Regressor	0.7725	0.5967	0.5861	0.4223
XGBRegressor	0.8005	0.6408	0.5419	0.3797
CatBoosting Regressor	0.7341	0.5388	0.5079	0.4784

TABLE 4. 21: Model Performance Test

We choose Linear regression as the final model because that model will get a training set r2 score is 0.82 which is 82% and a testing set r2 score is 0.82 which is also 82%.

4.7 Model Selection

This is used to select the best model of all of the regression algorithms.
In linear regression, we got 88.03 accuracy in all of the regression models that's why this model was the appropriate model to use for prediction.

4.7.1 Results for research question four

The specific objective two of this study was to identify how valid the developed model is for application in measuring the risk of incomplete grading in higher education. From this literature, we analyzed the data using different models which included Linear Regression, Lasso, K-Neighbors Regressor, Decision Tree, Random Forest Regressor, Gradient Boosting, XGBRegressor, CatBoosting Regressor, AdaBoost Regressor. In order to determine the appropriate model to use, we compared the model's performance training set and performance test set.

The validity of the model was determined by the accuracy in performance training set and performance testing set and we identified Linear regression as the appropriate model since it gave an accuracy of R2 score of 88.03 percent compared to the other models.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

The following section explores the conclusions derived from the model developed during the data collection and analysis of information from the selected Kenyan private and public universities. Furthermore, contribution of the study and recommendations by the researcher will be provided in this chapter.

The study developed a regression model to forecast the possibility of incomplete grading at institutes of higher learning. For students, educators, and institutions, incomplete grading—where ratings for student assessments are omitted or incorrect—presents substantial difficulties. The goal of this study was to create a prediction model that would help determine the causes of incomplete grading and its effects.

The model, which included factors linked to student assessments, course effects factors, instructor factors and even more so the student factors, was built using historical data collected through online questionnaires. The model successfully predicted the risk of incomplete grading by analyzing these variables. The study also looked at the effects of incomplete grading, such as how they can affect student welfare, academic achievement, and institutional responsibility.

The results emphasized the value of accurate data in higher education and the possibility of predictive modeling to enhance educational evaluation procedures. Institutions may improve the quality of instruction, support student achievement, and guarantee accountability in the assessment process by recognizing and addressing the risk of incomplete grading.

5.2 Conclusions

The complexity of this enduring problem has been clarified through studies on estimating the likelihood of incomplete grading in higher education institutions. Key findings from this research include:

Incomplete Grading Risk Factors: Student assessment challenges such as missed assessment followed by failure to communicate, Mode of assessment such as ignoring group assessments, data entry mistakes made by instructors, timely feedback to students once the assessment is done, challenges with the course content, selecting a course that aligns with your qualification is a big factor and technological challenges are primary factors contributing to incomplete grading. Institutions must address these factors to improve data accuracy.

Consequences of Incomplete Grading: Incomplete grading adversely affects student well-being, academic performance, increase in attrition rates and institutional accountability. The consequences are multifaceted and underscore the urgency of addressing this issue.

Predictive Model Effectiveness: The regression model developed in this study effectively predicts the risk of incomplete grading. By utilizing historical data and relevant variables, institutions can proactively identify and mitigate this risk.

Data-Driven Decision-Making: The model encourages data-driven decision-making in higher education. Institutions should leverage the insights provided by the model to enhance data accuracy and the overall educational experience.

The respondents drawn from the education sector were of the view that a better management of missing marks is essential for a sustainable prediction of incomplete grading when linear regression was being used. The study established that all learner assessment activities should be managed by the developed model of predicting incomplete grading. Moreover, through multiple regressions it is clear student factors has a positive relationship with sustainable chance of predicting incomplete grading while in the higher education sector. Subsequently, organizations management in higher education should dedicate more on providing more support and policies responsible for student factors. A functional system for managing missing marks requires transparency between the learner and the organization to ensure student factors do not lead to incomplete grades.

On the other hand, the respondents felt that organization policies were not clearly articulated when addressing issues related to incomplete grading in the higher education sector. Most organizations have not documented clearly the policies that guide on how to address issues with incomplete grading. Documentation of policies is vital in any organization as it ensures stakeholders have guidelines for particular situations they may find themselves in during implementation of their roles. Clear illustration of organization policies is preferred by the learner who use their personal time to undertake assessments. Moreover, the research established that having a clear and structured policy has a better way of managing issues with incomplete grading.

Respondents felt that the existing instructor factors was instrumental in how arising issues were addressed by the organization. Moreover, the respondents admitted that make up assessments should always be provided as this can aid in managing the risk of incomplete grades. It was not clear whether financial resources were a bigger factor in predicting the risk of incomplete grading as most government students are sponsored by the state. A few factored in that fee payment can subsequently lead to risk of having an incomplete grade.

Finally, course factors on Kenyan was identified as being essential for predicting the risk of incomplete grades among the learners. The respondents were of the opinion that some course and lessons are challenging to understand and this was important in developing a sustainable model to predict the risk of incomplete grading. Learner training and education heightens the knowledge held by the respondents when pursuing the course. Moreover, leaners need to be provided with course materials on a specific course, this can limit the risk of the learner having an incomplete grade. Subsequently, training resources for such activities can heighten the level of managing incomplete grades in the higher education sector.

5.3 Contributions of the Study

One of the primary contributions of this research is the development of a predictive regression model. This model provides higher education institutions with a practical tool to anticipate and assess the risk of incomplete grading. It offers a proactive approach to addressing data accuracy issues, helping institutions better manage their assessment processes.

The model encourages a shift towards data-driven decision-making in higher education. By identifying key risk factors for incomplete grading, institutions can make more informed choices regarding data management and grading processes. This approach enhances the overall quality and efficiency of educational assessments.

The research contributes to a more comprehensive understanding of the complex issue of incomplete grading. By examining various factors contributing to incomplete grading, such as grading errors, data entry issues, and technological challenges, it provides a holistic view of the problem, which is essential for developing effective solutions.

This study delves into the consequences of incomplete grading, including its impact on student well-being, academic performance, and institutional accountability. This assessment helps educational stakeholders recognize the real-world implications of incomplete grading and underscores the urgency of addressing this issue.

The research offers practical recommendations for higher education institutions. These recommendations guide institutions in enhancing data quality, improving data security, providing student support services, and developing clear grading policies. They serve as a roadmap for institutions looking to address the issue effectively.

By mitigating the risk of incomplete grading, the research contributes to promoting equity in education. It ensures that all students receive accurate and complete assessment records, minimizing disparities in academic outcomes. Moreover, it enhances institutional accountability by advocating for standardized grading practices.

The study sets the stage for further research in the field. By demonstrating the effectiveness of a predictive model, it encourages ongoing research to refine the model and explore additional variables that may contribute to incomplete grading. This paves the way for continuous improvement in data management practices.

The use of linear regression in predicting the risk of incomplete grades in the education sector is more pronounced as most higher education universities are experiencing the issues of missing marks. Subsequently, learners find the importance of class attendance and assessment attempts as vital and convenient for their course work. The current study adds knowledge regarding the prediction of incomplete grading in higher education by use of linear regression. Subsequently, the findings of this research augment the need for improved student behavior towards learning, better instructor support towards the student using programs such as MUSES. It is imperative to gather knowledge on how best to employ linear regression to predict the risk of incomplete grading while protecting the organization. The sentiments gathered by this research is mostly from the students who are the most populous demographic within the higher education sector. Subsequently, the knowledge will aid top management in better decision making regarding the management of incomplete grading. Incomplete grading such as missing marks will continue to evolve and it thus essential to use available resources to gather knowledge on how best manage the issues within organizations in the digital era.

This research offers a data-driven approach to addressing the issue of incomplete grading in higher education. It not only contributes a practical predictive model but also provides a comprehensive understanding of the problem, its consequences, and actionable recommendations for institutions. By doing so, it aims to improve data accuracy, promote equitable education, and enhance the overall educational experience for students in higher education institutions.

5.4 Recommendations for Future Research

Based on the findings of this study, the following recommendations are made to address the issue of missing marks in Kenyan universities:

Data Quality Improvement: Institutions should prioritize the improvement of data quality by addressing grading errors, implementing data entry best practices, and ensuring the reliability of technological systems.

Educator Training: Comprehensive training programs for educators on grading procedures and the use of digital grading systems can help reduce errors and improve data accuracy. Provide training and support to examiners and administrative staff to ensure they understand the grading system and procedures for recording grades. Foster a culture of academic integrity and ethical behavior among students and faculty members. By implementing these recommendations, universities can prevent and effectively address the issue of missing marks, thereby improving the overall quality and fairness of the academic experience for students.

Data Security Measures: Enhance data security and privacy measures to safeguard against technological challenges and data breaches, ensuring the integrity of student records. Enhance the technology infrastructure for grading and ensure that the systems are reliable and secure.

Student Support Services: Develop support mechanisms for students affected by incomplete grading, including academic counseling and guidance to mitigate the psychological impact. Improve communication between lecturers and examiners to ensure accuracy in grading and prevent missing marks. Provide support and counseling services for affected students to mitigate the negative impacts of missing marks on their mental health and academic performance.

Student Training: Increase awareness among students and faculty members about the importance of accurate grading and the consequences of academic misconduct.

Policy Development: Institutional policies should be updated to include clear grading guidelines and standardized practices, reducing the risk of incomplete grading. Establish clear policies and procedures for addressing missing marks, including prompt communication with affected students and the provision of clear information on the steps that will be taken to address the issue. Conduct regular audits of the grading system to identify and address any issues or errors.

Continued Research: Encourage ongoing research to refine the predictive model and explore new variables contributing to incomplete grading.

The prelevalance of missing marks in higher education sector where students face the risk of incomplete grading is a more important issue that needs to be handled with at most care as it can have a significant effect on the progress of the student and the need to protect the image of the university. It is imperative that such access is secure, convenient to the student and a well-managed system by organization is in place. The organization should develop a more robust model that can predict the risk of incomplete grading using more variables.

Institutions should ensure that the issues related to missing marks are well documented and made available to the students within the organization. Subsequently, the linear regression recommended by the researcher is one that lays focus on student factors that can lead to the prediction of incomplete grading in higher education, enhanced instructor factors and course effects.

REFERENCES

017-2018-university-statistics-report-approved-doc&id=18:universities-data-0-3&Itemid=187

Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. Computational Intelligence and Neuroscience, 2022.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 17, 1-21.

Ashfaq, U., Booma, P. M., & Mafas, R. (2020). Managing student performance: A predictive analytics using imbalanced data. International Journal of Recent Technology and Engineering, 8(6), 6.

Bakri, A. A., Wandanaya, A. B., Violin, V., & Fauzan, T. R. (2023). The Application of UTAUT Modified Model to Analyze the Customers Use Behavior of Shopee Paylater. Jurnal Sistim Informasi dan Teknologi, 96-101.

Bostedt, G., & Boström, L. (2022). Student Conceptions of Motivation to Study Revealed Through Phenomenography: Middle School and High School Students. In ECER 2022-The European Conference on Educational Research.

Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. Sustainability, 11(10), 2833.

Chui, K. T., Fung, D. C. L., Lytras, M. D., & Lam, T. M. (2020). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. Computers in Human Behavior, 107, 105584.

Commission for University Education (CUE). (2019). 2017-2018 University Statistics Report.

Damuluri, S., Islam, K., Ahmadi, P., & Qureshi, N. S. (2020). Analyzing navigational data and predicting student grades using support vector machine. Emerging Science Journal, 4(4), 243-252.

Deng, Y., Cherian, J., Khan, N. U. N., Kumari, K., Sial, M. S., Comite, U., ... & Popp, J. (2022). Family and academic stress and their impact on students' depression level and academic performance. Frontiers in psychiatry, 13, 869337.

Deveci, C. Ç., & Ersen, R. K. (2022). Predictors of Academic Achievement at Different Levels of Socioeconomic Status. International Education Studies, 15(3), 39-52.

Feng, T., Jia, X., Pappas, L., Zheng, X., Shao, T., Sun, L., ... & Ma, Y. (2022). Academic performance and the link with depressive symptoms among rural Han and minority Chinese adolescents. International Journal of Environmental Research and Public Health, 19(10), 6026.

Goss, H. (2022). Student learning outcomes assessment in higher education and in academic libraries: A review of the literature. The Journal of Academic Librarianship, 48(2), 102485.

Guay, F. (2022). Applying self-determination theory to education: Regulations types, psychological needs, and autonomy supporting behaviors. Canadian Journal of School Psychology, 37(1), 75-92.

Hamilton, L. S., & Ercikan, K. (2022). COVID-19 and US schools: Using data to understand and mitigate inequities in instruction and learning. Primary and secondary education during Covid-19: Disruptions to educational opportunity during a pandemic, 327-351.

Hossain, S. F. A., Xi, Z., Nurunnabi, M., & Anwar, B. (2022). Sustainable academic performance in higher education: a mixed method approach. Interactive Learning Environments, 30(4), 707-720.

Hossain, S. F. A., Xi, Z., Nurunnabi, M., & Anwar, B. (2022). Sustainable academic performance in higher education: a mixed method approach. Interactive Learning Environments, 30(4), 707-720.

Jongile, S. (2022, December). Contextualisation of Predictor Variables for Students' At-Risk of Dropping Out of University. In EdMedia+ Innovate Learning Online 2022 (pp. 337-357). Association for the Advancement of Computing in Education (AACE).

Kamau, E. (2019). The Effect of Missing Marks on Students' Academic Performance: A Case Study of Moi University. Journal of Educational Policy and Entrepreneurial Research, 6(3), 69-77.

Kamble, V. B., & Deshmukh, S. N. Comparative Analysis Of Standard Error Using Imputation Method. International Journal of Innovations in Engineering Research and Technology, 1-6.

Kanetaki, Z., Stergiou, C., Bekas, G., Jacques, S., Troussas, C., Sgouropoulou, C., & Ouahabi, A. (2022). Grade prediction modeling in hybrid learning environments for sustainable engineering education. Sustainability, 14(9), 5205.

Karakose, T., Polat, H., Yirci, R., Tülübaş, T., Papadakis, S., Ozdemir, T. Y., & Demirkol, M. (2023). Assessment of the relationships between prospective mathematics teachers' classroom management anxiety, academic self-efficacy beliefs, academic amotivation and attitudes toward the teaching profession using structural equation modelling. Mathematics, 11(2), 449.

Kathambi, E. (2019). Factors Contributing to Missing Marks in Kenyan Universities: A Case Study of Jomo Kenyatta University of Agriculture and Technology. Journal of Education and Practice, 10(3), 1-10

Keane, T., Linden, T., Hernandez-Martinez, P., & Molnar, A. (2022). University students' experiences and reflections of technology in their transition to online learning during the global pandemic. Education Sciences, 12(7), 453.

Leaper, C. (2023). Gender and social-cognitive development.

Lund, B. (2023). The questionnaire method in systems research: An overview of sample sizes, response rates and statistical approaches utilized in studies. VINE Journal of Information and Knowledge Management Systems, 53(1), 1-10.

Marlina, E., Tjahjadi, B., & Ningsih, S. (2021). Factors affecting student performance in e-learning: A case study of higher educational institutions in Indonesia. The Journal of Asian Finance, Economics and Business (JAFEB), 8(4), 993-1001.

Mishra, P. M., & Saboowala, R. Imputing Missing Values for Data Samples In Educational Empirical Research Using Machine Learning Techniques. XAVIERIAN JOURNAL OF EDUCATIONAL PRACTICE, 23.

Mugenda, O. M., & Mugenda, A. G. (2003). Research methods: Quantitative & qualitative approaches (Vol. 2, No. 2). Nairobi: Acts press.

Muryan Awaludin, M. (2022). Optimization of Naïve Bayes Algorithm Parameters for Student Graduation Prediction at Universitas Dirgantara Marsekal Suryadarma. Journal of Information System, informatic and Computing, 6(1), 91-106.

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237.

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237.

Pedagogies, 21, 1-34.

Poudyal, S., Mohammadi-Aragh, M. J., & Ball, J. E. (2022). Prediction of student academic performance using a hybrid 2D CNN model. Electronics, 11(7), 1005.

Rahman, M. A., Novitasari, D., Handrianto, C., & Rasool, S. (2022). Challenges in online learning assessment during the covid-19 pandemic. Kolokium Jurnal Pendidikan Luar Sekolah, 10(1), 15-25.

Rawlusyk, P. E. (2018). Assessment in higher education and student learning. Journal of Instructional

Ruiz-Jimenez, M. C., Martinez-Jimenez, R., Liceran-Gutierrez, A., & Garcia-Marti, E. (2022). Students' attitude: Key to understanding the improvement of their academic RESULTS in a flipped classroom environment. The International Journal of Management Education, 20(2), 100635.

Sudais, M., Safwan, M., Khalid, M. A., & Ahmed, S. (2022). Students' Academic Performance Prediction Model Using Machine Learning.

Trivedi, S. (2022). Improving students' retention using machine learning: Impacts and implications. ScienceOpen Preprints.

Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behavior, 104, 106189.

Wright, N. A., & Arora, P. (2022). A for effort: Incomplete information and college students' academic performance. Economics of education review, 88, 102238.

Appendix 1: Research Schedule

TABLE 4. 22: Research schedule

2023	JAN	FEB	MAR	APR	MAY	JUNE	JUL	AUG	SEPT	OCT
Concept Ideation										
Draft Research q.										
Writing Proposal										
Review of lit.										
Proposal Pres.										
Data Collection										
Model formulation										
Data analysis										
Model validation										
Work compilation										
Final defense										
Document submission										

Appendix 2: Resources and Budget

Table 4.23. Proposed Budget

No.	Item Name	Unit	Price	Total
1.	Computer System	1	33000	33000
2.	Internet connectivity	5	5000	25000
3.	Travel cost	30	1500	4500
4.	Data collection	1	1500	1500
5.	Final Dissertation	500	30 per page	15000
6.	Flash Drive	1	1500	1500
7.	Miscellaneous expenses	1		5500
8.	Softwares for analysis	2	7000	14000
9.	Total			100,000

Appendix 3: Questionnaire

You are kindly requested to complete the attached questionnaire. Please, note that all the informationgiven shall be purely used for academic purposes and shall be treated as confidential. Thank you for taking your time to complete the questionnaire and for your cooperation.

PART A. GENERAL INFORMATION

1. Gender

   Male ☐ Female ☐ Other ☐

2. Age

   Under 20 years ☐ 20-30 years ☐ 31-40 years ☐ over 41 years ☐

3. County of origin

   Nairobi ☐ Other ☐

4. Are you a self-sponsored student?

   Yes ☐ No ☐

5. How long have you been studying?

   Less than one month ☐

   More than a month less than 6 months ☐

   More than one year ☐

   Less than 3 years ☐

   More than 3 years ☐

6. Which one of the following courses do you take?

   - Information Technology ☐
   - Business Management ☐
   - Education ☐
   - Commerce ☐
   - Economics & Statistics ☐
   - Data Science ☐

- Medicine
- Psychology
- Other ……………………………

PART B. EFFECT OF INSTRUCTOR FACTORS ON STUDENT INCOMPLETE GRADING

Please tick (√) to indicate the extent to which you agree or disagree with the following statements effect of customer attitude on product brand image. Use the scale:

Note: 1= Not at all (NAA) 2 = Small extent (SE) 3 = Moderate extent (ME)

4 = Large extent (LE) 5 = Very large extent (VLE)

	INSTRUCTOR FACTORS	**NAA**	**SE**	**ME**	**LE**	**VLE**
1	I believe that the lecturer plans his/her lessons well					
2	I believe this course is well prepared					
3	I am aware that this course has adequate resources to support learning provided by the lecturer					
4	I believe the lecturer lessons are easy to understand					
5	I believe the lecturer is punctual					
6	The instructor has adequate knowledge about this course					
7	I am quite satisfied with his/her teaching					
8	I am confident assignments are checked and feedback is given within an acceptable period of time					
9	The instructor often issues assignments on virtual campus					
10	The instructor often issues cat assessment on virtual campus					
11	I am confident cat assessments are checked and feedback is given within an acceptable period of time					
12	I enjoy the teaching approach of the instructor					
13	I believe the lecturer issues all assignments as required by institution					

14	The instructor is easily approachable in preferable times					
15	I believe the instructor frequently misses classes					
16	The instructor often issues the course outline before session starts					

PART C. EFFECT OF STUDENT FACTORS ON INCOMPLETE GRADING

	STUDENT FACTORS	**NAA**	**SE**	**ME**	**LE**	**VLE**
1	I often attempt all assignments issued by the instructor					
2	I prefer group work assessment					
3	I believe i have a good class attendance record					
4	I often communicate in case I miss an assessment					
5	I engage my lecturer frequently					
6	I often experience fee issues most time with my study					
7	Most units being taught has adequate exercises for practice					
8	The instructor issues in class assessments					
9	The instructor gives scores for attempts made during class session exercises					
10	The lesson is taught for the time period indicated					
11	Make up is issued in case i miss an assessment					
12	The lecturer always communicates the urgency of missing coursework					

PART D. EFFECT OF COURSE FACTORS TO STUDENT INCOMPLETE GRADING

	COURSE EFFECT	NAA	SE	ME	LE	VLE
1	I have a challenge understanding the course unit					
2	I believe this course is significant for my major					
3	My instructor has adequate knowledge about this course					
4	The course outline is issued before the session starts					
5	The course session plans are well prepared					

PART E. RISK OF INCOMPLETE GRADING

	EFFECT	NAA	SE	ME	LE	VLE
1	I have considered dropping out of the university due to missing marks					
2	I have been forced to repeat the unit due to missing marks					
3	I believe cases of missing marks has led to delayed graduations					
4	Missing marks complaints are not being addressed properly in the organization					
5	I believe the organization is doing enough to handle issues related to missing marks					
6	I believe there are proper systems in place to manage missing marks					

PART F. STUDENT ASSESSMENT FACTORS

	EFFECT	NAA	SE	ME	LE	VLE
1	All assessments are issued by the instructor					
2	Assessment is attempted is submitted in good time					
3	I normally attempt all assessments					
4	Assessment feedback is timely					
5	I believe I normally attempt all assignments and cats					
6	I believe all assessments should be conducted through virtual campus					
7	I prefer assessments to be done in groups					