AN EDUCATIONAL DATA MINING MODEL FOR PROMOTING SELF-REGULATED LEARNING ON LEARNING MANAGEMENT SYSTEMS

ARAKA N. ERIC

BSC (Kenyatta University), MSC (University of Nairobi)

J57/25721/2018

A thesis submitted in partial Fulfillment of the Requirements for the Award of the Degree of Doctor of Philosophy (Computer Science) in the School of Engineering and Architecture of Kenyatta University

March, 2023

DECLARATION

I declare that this thesis is my original research work and has not been presented in any other university or institution for consideration for any certification. The thesis has been complemented by referenced sources duly acknowledged. Where text, graphics, pictures, figures, or tables have been borrowed from other sources, including the internet, these are specifically accredited and references cited using the current APA system and in accordance with anti-plagiarism regulations.

Signature.................................... Date...22/03/2023...

Supervisors' declaration: This thesis has been submitted for appraisal with our/my approval as University Supervisor(s):

Signature.................................... Date...22/03/2023...

Dr. Elizaphan Maina
Department of Computing & Information Technology
School of Engineering & Technology
Kenyatta University

Signature.................................... Date...22/03/2023...

Prof. Robert Oboko
School of Computing and Informatics
University of Nairobi

Signature.................................... Date...22/03/2023...

Dr. Rhoda Gitonga
Department of Computing & Information Technology
School of Engineering & Technology
Kenyatta University

DEDICATION

This research is dedicated to my wife, family members, and colleagues from whom I received overwhelming support in my study.

ACKNOWLEDGEMENTS

First, I would like to thank God for giving me good health and supplying me with the wisdom that has helped me carry out this research.

Secondly, I pass my appreciation to my supervisors Dr. Elizaphan Maina, Prof. Robert Oboko, and Dr. Rhoda Gitonga for their guidance, support, and inspiration throughout this research. Their knowledge and insights into this research gave me the courage to soldier on.

Thirdly, I wish to thank the National Research Fund (NRF), Kenya for supporting this research. The funding helped me carry out this research without any financial constraints. I also thank the entire NRF research team comprised of the faculty from the University of Nairobi, The Cooperative University of Kenya, and Kenyatta University; Prof. Peter Wagacha, Prof. Kennedy Waweru, and Prof. John Kihoro especially for his statistical insights throughout the entire research process.

I thank all my friends and colleagues for their moral support accorded to me during my research and thesis writing. I am especially grateful to Dr. Elizabeth Mwaniki for her time and interest in my studies. Her insightful feedback helped me to see where I could not. I am also thankful for my colleagues at the Computing & Information Technology department of Kenyatta University for their moral support and advice throughout the research process.

Finally, I am very grateful to my wife, Purity Njoki, for putting up with me from the start to the completion of my doctorate research. She is always a source of encouragement for me to pursue what I have always desired to accomplish.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIHS	Adaptive & Intelligent Hypermedia Systems
EDM	Educational Data Mining
ITS	Intelligent Tutoring Systems
LA	Learner Analytics
LASSI	Learning and Study Strategies Inventory
LMS	Learning Management Systems
MOOC	Massive Open Online Course
MSLQ	Motivated Strategies for Learning Questionnaire
OSLQ	Online Self-Regulated Learning Questionnaire
PLE	Personal Learning environments
PLSA	Personalized Learning Student Actions
SJT	Situational Judgment Tests
SRES	Student Relationship Engagement System
SRL	Self-Regulated Learning
SRLIS	Interviews such as Self-Regulated learning Interview Schedule
SRMC	Self-Regulation Measure for Computer-based Learning
IDP	Integrated Design Process
SSRL	Socially Shared Regulated Learning
MASRL	Metacognitive and Affective Model of Self-Regulated Learning

ABSTRACT

Online learning environments such as Learning Management Systems (LMS) utilized by institutions of higher learning to deliver open and distance education have different features that may not adequately provide individualized support to learners. Online learners experience inadequate instructor support as many students are enrolling in fully online or blended courses. Therefore, the success of online learning depends on the learner's ability to take control of their learning process as defined by the Self-Regulated Learning (SRL) theory. Since online learners have no restricted time to go online and learn, the existing freedom requires students who can take charge and control their learning. Moreover, existing models for supporting SRL do not link SRL in educational psychology and SRL in LMS. In view of this, it is necessary to have LMS that support SRL by providing targeted interventions based on students' online learning activities data. This study explored the use of Educational Data Mining (EDM) techniques to support students' SRL skills on LMS. To accomplish this, the research applied a mixed-methods approach that involved three phases. First, qualitative and quantitative methods were used in problem identification. Qualitative research was used to explore the current methods used to measure and promote SRL in online learning environments. Quantitative research was used to conduct a pre-study to establish how LMS features are utilized in teaching and learning in higher institutions of learning in Kenya. Secondly, Integrated Design Process was used to design clustering and classification algorithms and integrated them into Moodle LMS to analyze students' learning activities data. Finally, true experiment research was used to establish the effectiveness of the developed EDM model in promoting SRL through sampled University students who enrolled in data science with python course for 12 weeks. This study, therefore, contributes an EDM model that contains algorithms derived from learning theories as applied in self-regulated learning. Results indicate that students were able to progress from poor self-regulators cluster level to good self-regulators and exemplary self-regulators. The results reveal that students were able to evolve from one cluster to another over time as a result of receiving EDM interventions during the online course. The findings demonstrate that providing external support through the use of EDM prompts leads to the growth of students' SRL skills. Moreover, this study reveals that it is possible to measure and promote SRL concurrently using EDM techniques. Future studies can focus on the evaluation of the EDM model under varied contextual conditions to examine the effect of SRL interventions on students' SRL skills and academic performance.

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter presents the background to the study, the statement of the problem, the purpose of the study, and the objectives. The research questions, justification, significance of the study, and the scope of the study are covered as well in this chapter.

1.2 Background of the Study

The advancements in technology-enhanced teaching and learning provide a continuum of educational environments from physical classroom interactions to asynchronous interactions in online learning. Consequently, Institutions of Higher Learning (IHLs) have continued to adopt blended and fully online courses. Many IHLs adopt online learning for fully online courses (asynchronous where students learn independently by accessing learning materials via LMS) or blended learning (where LMS is used to support learners in synchronous class) to resolve the challenge of increased enrollments for higher education in Universities and colleges (Hadullo et al., 2018; Makokha & Mutisya, 2016; Otieno, 2016). Additionally, IHLs adopt online learning to compensate for the limited number of teaching staff and since it also requires fewer infrastructural facilities (Arkorful, 2014).

Despite the advancements in technology, recent studies reveal that there is limited instructional support offered to learners undertaking online studies (Araka et al., 2021). The large number of students enrolling in online courses has made it difficult to offer individualized instructional support due to the limited number of online instructors (Romero & Ventura, 2017). Compared to their counterparts who attend physical classroom teaching and are confined to certain periods while they are regularly guided by instructors, online learners are not restricted in managing their schedules in terms of the time of study and study duration. Consequently, the success and of online learning rest on the learners' ability to self-regulate.

Self-Regulated Learning is a theory in educational psychology that describes the process through which students plan, monitor, and evaluate their learning by employing appropriate strategies such as time management, effort regulation, help-seeking, motivation, and cognition (Garcia et al., 2018; Pintrich & Zusho, 2002; Zimmerman, 1986). Self-regulated learners can manage their learning and therefore assume an active role in achieving their academic goals (Zimmerman, 1990). According to Pintrich (2000), SRL is a constructive process that enables learners to have control of their learning by playing an active role in the learning process while being guided by their motivation, behavior, and metacognition.

With the increased focus on e-learning, students must possess the ability to use digital tools and online learning environments' features effectively. Literature indicates that those students who actively employ SRL skills during learning perform better than those who do not. Students who take an active role in e-learning have been found to perform better compared to those who possess low levels of SRL skills (Cavalcanti et al., 2018; Cicchinelli et al., 2018). Additionally, it has been established that students' academic achievement can be determined and predicted by measuring the level of engagement in an LMS (Hashemyolia et al., 2014).

The measurement of SRL has undergone two major phases (Panadero et al., 2016); the first phase is when SRL was regarded as trait-based and individually inclined and was measured based on learner characteristics without taking into context the learning environment. The tools that were developed during this era were oriented towards measuring SRL based on student traits such as self-reports. The advancement of technology and research has led to the second phase where SRL measurement techniques were domain-based. During this era, SRL was measured during the learning process while taking into consideration of the actual learning environment such as the classroom or online learning environment. The usage of self-report tools to measure SRL has some limitations. First, there is a tendency for learners to overestimate the use of their SRL skills. Self-report tools are user-oriented. Hence learners will only report what they perceive about themselves. Secondly, the self-report measures are deployed without considering the actual learning environment and therefore measure SRL before or after a learning period (Roth et al., 2016a; Siadaty, 2016; Winne & Perry, 2000). Researchers argue that SRL can only be measured in the context of an actual learning environment where it occurs (Lee, 2008). Whereas the adoption of technology in teaching and learning requires change or adoption of improved methods of measuring SRL, there has been the continued use of traditional tools that were designed to measure SRL in face-to-face classroom settings (Broadbent & Poon, 2015).

As technology advanced, online SRL measurement tools were developed. Online tools are designed to measure SRL during a learning episode. For instance, log data generated from Massive Open Online Courses (MOOC) and LMS can be analyzed to infer SRL skills for each learner.

With online measures, SRL is viewed as process-oriented and therefore can be measured as a series of events (Winne & Perry, 2000). The advantage of online tools over self-report tools is that online tools (collecting and inferring SRL strategies from log data) are unobtrusive since SRL is measured in the background and therefore does not affect the engagement and behaviors of the students (Schraw, 2010).

The emergence of the Educational Data Mining (EDM) field introduces new dimensions where trace data from educational systems is collected, explored, analyzed, and inferred to understand learners and improve learning environments. For example, Arnold & Pistilli (2012) in their study, employed EDM on both online (log data from online learning environments) and offline data (data collected from self-report tools such as questionnaires) to deliver visualized reports to teachers about students' SRL levels and make it easy to monitor SRL for learners and submit feedback to them. EDM is also being used to measure students' self-regulatory skills in MOOCs and LMS and present visualized reports to learners through dashboards (Lee & Recker, 2017; Nussbaumer, Hillemann, & Albert, 2015; Nussbaumer, Hillemann, Gütl, et al., 2015; A. Rodriguez et al., 2014). The presentation of analyzed reports to instructors who then presents a summary to learners, however, presents a challenge of human limitation; the ability to handle the increased number of students and that human judgment can be prone to error. Moreover, there have been calls for researchers to develop tools that can provide both SRL measurements and interventions at the same time for e-learning students (Panadero et al., 2016). The existing approaches, therefore, demonstrate lack of robust EDM models that perform EDM to achieve SRL. For instance, the exiting models do not link SRL in educational psychology and SRL in LMS. Additionally the existing approaches (models) of SRL promotion fail to integrate algorithms to achieve robust identification of undesirable learner behaviors and classify them into clusters to allow for provision of targeted interventions in real-time as students learn on LMS.

Since online learning environments such as LMSs and MOOCs place the learner at the center of the learning process, they should provide tools for measuring and promoting SRL as e-learning is a complex process that is dependent on learning environments and requires that learners regulate their learning (Azevedo & Witherspoon, 2009). This necessitates the provision of SRL interventions to learners to support them in monitoring their learning since no one can be able to self-regulate when left on their own individually or collectively. Furthermore, it will be necessary for online learning environments to provide tools that can measure SRL skills without learners being aware. Moreover, it will make it easier to identify those who need interventions and then decide the type of interventions they need and to what extent (Järvelä & Hadwin, 2013; Lodge et al., 2019).

1.3 Problem Statement

Since the emergence of self-regulated learning over three decades ago, the prevalent mode of learning and teaching has been the traditional face-to-face classroom setup. In this mode, self-report tools such tools as the Motivated Strategies for Learning Questionnaire (MSLQ), Learning and Study Strategies Inventory (LASSI), and the Self-Regulated Learning Interview Schedule (SRLIS) are used to measure SRL (Roth et al., 2016a). The self-report tools measure SRL before or after a learning episode. (Winne & Perry, 2000) observes that self-report tools face major drawbacks; students can overestimate their SRL skills hence the outcome is dependent on learners' perception of their SRL skills. Also, the self-report tools are deployed outside a learning environment (Roth et al., 2016a; Winne & Perry, 2000). Lee (2008) argues that when measuring SRL, there is a need to take into consideration the learning environment within which learning occurs. The use of self-report tools does not allow for the provision of interventions for learners. Thus, little knowledge exist about how to provide SRL interventions based on SRL evaluation results based on the offline tools. The adoption of online learning as a continuum to physical classrooms introduced a different perspective to the definition, measurement of SRL, and the provision of SRL interventions. Consequently, the concept of viewing SRL as trait-based changed to a process oriented that is measured as a series of events. This gave rise to the use of online tools for measuring SRL through log traces. However, most studies rely on the use of data generated from classroom observations, laboratory studies, self-test scores, self-reporting techniques, and surveys to measure the level at which learners employ SRL in online learning environments (Araka et al., 2020; Kizilcec et al., 2017; Saks & Leijen, 2014). Presently, EDM is being applied in the measurement of SRL in online learning environments to predict students' academic performance (Cavalcanti et al., 2018; Cicchinelli et al., 2018; Hashemyolia et al., 2014). There have also been calls for the development of tools that can not only measure SRL on e-learning platforms but also develop interventions to promote SRL in e-learning environments (Alario-Hoyos et al., 2017; Panadero et al., 2016; Terras & Ramsay, 2015). Moreover, the review of current research indicates a lack of a model to be used to measure SRL through the use of EDM tools. Existing studies emphasize on measuring learners' levels of SRL but fail to explore the effect of SRL interventions on the growth of students' self-regulatory skills. For instance, it will be interesting to establish if learners with low-level SRL skills can be motivated by SRL interventions to achieve higher levels of SRL skills. Literature indicates many studies only focus on using EDM to measure SRL rather than promote SRL that will help learners to improve their SRL skills and even academic performance (Araka et al., 2020; Viberg et al., 2020). Despite the increased adoption of online courses, institutions of higher learning have continued to rely on instructors to provide support to online learners. However, instructors may not be able to handle a large number of enrolled students (Muuro et al., 2014). There is a lack of empirical evidence on the effect of SRL interventions on students learning outcomes and how the SRL interventions may help enhance students' SRL skills in online courses (Broadbent et al., 2020). There is a need therefore to investigate how EDM can provide interventions based on data collected and analyzed from online learning environments. Consequently, there is a need for empirical research to establish the effectiveness of using EDM tools compared to human interventions in scaffolding SRL in institutions of higher learning. Without SRL interventions, online learners lack the motivation to learn and are less expected to take control of their learning hence leading to student dropout, reduced retention rates, and prolonged stay at university than expected (Karaoğlan Yılmaz et al., 2018). This study, therefore, endeavored to investigate how support for online learners can be provided to facilitate the development and growth of self-regulated learning skills using EDM techniques.

1.3.1 Research Purpose

The purpose of this research was to explore how EDM techniques can be used to promote Self-Regulated Learning on Learning Management Systems.

1.3.2 Research Objectives

This research was guided by the following objectives:

1. To investigate the methods currently being used to measure and promote SRL in online learning environments.
2. To investigate the current status of utilization of LMS features in promoting SRL in online learning.
3. To develop and integrate into Moodle LMS an intelligent model that applies EDM to promote SRL strategies on LMSs.
4. To evaluate the EDM model for its effectiveness in promoting SRL strategies on LMS.

5. To establish students' perceived usefulness of the EDM model in promoting SRL on LMS.

1.3.3 Research Questions

This research was guided by the following research questions:

1. What methods are currently being used to measure and promote SRL strategies in online learning environments?
2. What is the current status of utilization of LMS features in promoting SRL strategies in online courses?
3. How can an EDM model be developed and integrated into Moodle LMS to promote SRL?
4. How effective is the EDM model in promoting SRL on LMS?
5. What is students' perceived usefulness of the EDM model in promoting SRL on LMS?

1.3.4 Research Hypotheses

To test for the effectiveness of the EDM model in promoting SRL on LMS, the following hypotheses were formulated:

1. Null Hypothesis: There is no significant difference between SRL strategies utilization by students in EDM-based intervention and instructor-based intervention groups.
2. Alternative Hypothesis: There is a significant difference between SRL strategies utilization by students in EDM-based intervention and instructor-based intervention groups.

Table 1.1 presents the mapping of research objectives and research questions to research methods that were utilized in this study.

Table 1.1: Mapping of Research Objectives and Research Questions to Research Methods

Objective	Question	Approach/Methodology
To investigate the methods currently being used to measure and promote SRL in online learning environments.	What methods are currently being used to measure and promote SRL strategies in online learning environments?	Document Analysis
To investigate the current status of utilization of LMS features in promoting SRL in online learning.	What is the current status of utilization of LMS features in promoting SRL strategies in online courses?	Quantitative research (Semi-structured survey)
To develop and integrate into Moodle LMS an intelligent model that applies EDM to promote SRL strategies on LMSs.	How can an EDM model be developed and integrated into Moodle LMS to promote SRL?	Integrated design process
To evaluate the model for its effectiveness in promoting SRL strategies on LMS.	How effective is the model in promoting SRL on LMS?	True experiment research
To establish students' perceived usefulness of the EDM model in promoting SRL on LMS?	What is students' perceived usefulness of the EDM model in promoting SRL on LMS?	Qualitative research (Structured survey and Semi-structured interview)

1.4 Justification

The main challenge with current technology-enhanced educational systems such as LMS is the lack of adequate learner support and guidance for interactive learning. This has been contributed by inadequate human support as many students are enrolling in e-learning platforms (Araka et al., 2021; Muuro et al., 2014; Muuro & Kihoro, 2017). The other challenge that has also been raised is the quality of teaching and learning where instructors use online learning platforms to disseminate learning materials with little attention given to the learning activities students engage in. The solution to these challenges lies in the provision of e-learning systems that possess in-built intelligence functionality that can determine the level of involvement in e-learning activities for each learner and be able to scaffold each learner accordingly.

This study employed Educational Data Mining techniques on educational data to integrate Moodle LMS with in-built intelligence that can offer adequate support and guidance to learners as far as Self-Regulated Learning is concerned. The result is to have technologically enhanced learning environments that are more efficient and effective compared to inadequate human interventions. The research contributes an EDM plugin for promoting SRL in the Moodle LMS. The study also examined the effect of offering SRL interventions based on educational data mining from educational environments compared to instructor-based interventions.

1.5 Scope

This study explored the use of educational data mining (EDM) techniques to support students' SRL skills on LMS. An EDM model, also referred in this thesis as an EDM plugin, was and developed. The model was integrated into Moodle LMS to offer metacognitive interventions to trigger students' thought processes and have them take charge of their learning. The viability and effectiveness of the EDM model was evaluated by conducting a comparative study using experimental group (students that used the EDM enabled LMS) and control group (students using traditional LMS). The study population was undergraduate students from institutions of higher learning in Kenya.

1.6 Significance of the study

This research intends to benefit Institutions of Higher Learning (IHLs) and e-learning researchers and practitioners. First, the IHLs in Kenya are increasingly adopting e-learning platforms to deliver learning materials. Most universities and colleges focus on one-size-fits-all instructor-centered learning rather than personalized student-centered learning. The implementation of EDM-based SRL interventions comes in handy for IHLs in providing individualized support and guidance to online students by incorporating into LMS a plugin to offer personalized learning features for active learning. This in the long run, is likely to lead to increased satisfaction and motivation for online learners and improved performance and success rates. Secondly, the findings from this research may benefit those engaged in technology-enhanced research and also enable e-learning practitioners to implement outcomes of the research in teaching and learning.

1.7 Limitations of the Study

First, in this study, the MSLQ self-report tool was used to measure SRL strategies for the students before and after the experiments. Researchers argue that self-report tools are biased and dependent on how students perceive themselves in terms of how they utilize SRL skills before and after learning. As a result, there were noticeable differences between the actual SRL strategies inferred from LMS logs and those reported by students via the MSLQ tool. Consequently, this study used the online approach of identifying the SRL skills from log data. However, this approach can only provide SRL measures at the end of a course. If this approach could be applied in future studies, it is likely to be difficult to measure SRL before the course since online tools only work on logs that are generated after an online course.

Secondly, this study demonstrates that automated interventions can be presented to learners on the go as they study within LMSs. However, it will be interesting to investigate the possibility of having instructors involved in the provision of automated support, especially when dealing with at most risky students. For example, future studies may consider scenarios where the analyzed information can first be presented to instructors and could decide and customize effective intervention messages that can be sent to the learners (Khosravi et al., 2021). According to Khosravi et al. (2021), such learners can be presented to the instructors who can then decide on the nature and effectiveness of the interventions to provide to the learner. Researchers have argued that artificial intelligence systems need to have instructors in the loop when offering EDM-based interventions. Future studies should therefore consider establishing the approaches of designing automated intervention systems that bring instructors in the know of what students receive and allow instructors to offer customized support, especially for high-risk students. This is likely to enhance social interaction between instructors and students (Seo et al., 2021). Future studies may therefore consider how such recommendations to instructors can be investigated and implemented while taking into consideration the large number of students enrolled in online learning.

Lastly, the intervention message triggers for student clustering and provision of intervention messages relied on Moodle LMS's APIs. The configuration of Moodle LMS APIs limited customization of the plugin on varying the frequency of intervention messages to students. The frequency of sending out SRL prompts should be differed depending on the cluster in which the students were placed in real-time. For instance, poor self-regulators in cluster 0 were should be prompted more compared to the students to exemplary self-regulators in cluster 3. Almost all the APIs in Moodle are configured in a way that, any choice of API affects all students enrolled in the course since they capture students' learning activities without any limitation. As a result, too many messages were sent to students in one learning episode since the APIs are configured in a way that captured more than one event. This made it difficult in limiting the frequency of sending out the interventions as well as varying the frequency based on the cluster in which a student is placed.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The adoption of technology-enhanced learning by most institutions of higher learning for Open and Distance Learning (ODEL) or blended learning coupled with increased demand for higher education has led to an increased number of students undertaking e-learning courses (Broadbent & Poon, 2015; Hashemyolia et al., 2014). However, literature reveals that there is inadequate support offered to learners who undertake their studies online. Researchers, educators, and psychologists opine that the success and satisfaction of online learners depend on the learners' ability to be in charge of their learning processes. Since self-regulatory skills are not inherent to learners, and there is no student, individually or collectively, who can develop self-regulation skills by themselves learning environments that enhance the development of such skills need to be designed (Cerezo et al., 2016; Dabbagh & Kitsantas, 2005). Although online learning environments such as LMSs and MOOCs contain an in-built set of tools and features, the evidence on how utilization of such features by learners effect of development self-regulatory skills remain scanty.

Online learners are not restricted in managing their schedules in terms of the time of study and duration. As a result, the success and effectiveness of online learning depend on the learner's ability to play an active role in the learning process. This kind of freedom and flexibility may be disadvantageous to online learners who experience little or no interactions with instructors. According to Vovides et al. (2007), for learners to develop and grow self-regulatory skills, they require contextual environments that enable them to plan, monitor, and evaluate their learning. Research studies indicate that students with high self-regulatory skills achieve better academic results compared to those with low self-regulatory skills. However, techniques that provide effective interventions to learners and the effect of such intervention on learners in improving self-regulatory skills and academic achievement has not been addressed. Additionally, there have been calls for the development of tools that can not only measure self-regulatory skills but also provide interventions to promote students' SRL skills for online courses (Alario-Hoyos et al., 2017; Panadero et al., 2016; Terras & Ramsay, 2015). This, therefore, necessitates an empirical study to investigate how EDM can be employed to measure SRL and support SRL interventions in the development of students' SRL skills.

This rest of chapter 2 is organized as follows;

Section 2.2 introduces the concept of Self-Regulated Learning theory and the conceptual frameworks in which SRL is grounded are discussed. Section 2.3 highlights the existing theories of learning and discusses the constructivist learning theory adopted in this study. Section 2.4 discusses the tools and techniques for measuring SRL in e-learning platforms and the challenges and research gaps associated with the existing tools. Section 2.5 discusses the techniques for promoting self-regulatory learning in online learning environments. Section 2.6 discusses Learning management systems. Section 2.7 highlights the potential of EDM algorithms in the measurement of SRL and provision of SRL interventions to learners. The models currently being employed in the measurement and promotion of SRL are also described. The EDM applications used for measuring learner behaviors in e-learning environments are also discussed. Section 2.8 presents the research gap analysis. Section 2.9 describes the conceptual framework that guided the design of experiments to validate the conceptual model. Section 2.10 presents the operationalization of the variables identified from the conceptual framework. Finally, section 2.11 gives the summary of the literature chapter.

2.2 Self-Regulated Learning

Self-Regulated Learning (SRL) is a theory that describes how and what learners should possess to be able to take control of the learning process. Learners who possess high self-regulatory skills can control and manage their learning by assuming active roles to achieve their academic goals (Zimmerman, 1990). The SRL theory involves the use of various strategies to control and monitor one’s aspects of metacognition, behavior, and motivation to achieve set goals for an online course. Students who possess high levels of self-regulatory skills learn faster and perform better than those with low levels of self-regulatory skills. The students also have more self-efficacy which makes them persist in learning because they believe they can be successful. (Kizilcec et al., 2017).

To effectively be able to provide learner support and guidance through SRL measurement and interventions, this study reviews the SRL strategies and how they should be manifested from LMS data and described by the existing SRL theoretical frameworks. According to (Zimmerman, 1990), the following are the main components of SRL;

1. ***Metacognition***: This component describes how learners set academic goals, plan how to achieve the goals, employ various strategies to achieve the set goals, and monitor the

learning process as a whole. Examples of self-regulation skills related to metacognition include one's orientation before the start of an assessment task such as quizzes and assignments, sourcing for relevant learning materials, integrating diverse viewpoints during a learning episode, and monitoring one's comprehension and progress in learning (Boekaerts & Cascallar, 2006). These examples demonstrate how learners build their cognitive skills and therefore become aware of themselves.

2. ***Motivation***: This component is one through which students can stay on course despite the obstacles encountered as they learn through motivation, students can come up with strategies that will help them navigate through the challenges encountered during the learning process. Students regulate their motivation by initiating learning activities and ensuring that they stay on course from the start, until the end of the learning activity. Students through motivation, learn how to initiate and maintain invested efforts to ensure successful completion of learning activities (Boekaerts & Cascallar, 2006).
3. ***Behavior***: Learner behaviors may dictate the effort committed to learning. Motivated students seek help from the learning community including fellow students and lecturers and engage in activities that help them continue learning by themselves. Help-seeking, one's persistence in handling difficult tasks, and invoking a positive mindset, attitude, and talk such as "let me keep trying" form part of the behavioral choices students have to make during the learning process to demonstrate their possession of self-regulated learning skills (Pintrich & Zusho, 2002).

Literature indicates that those students who actively employ SRL skills during learning perform better compared to those who possess low levels of SRL skills. Students' academic achievement can be determined and predicted by measuring the level of engagement in an LMS (Barnard et al., 2010; Cavalcanti et al., 2018; Cicchinelli et al., 2018; Hashemyolia et al., 2014). Studies indicate that self-regulation in learning can be used to predict students' academic performance. For example, in a systematic review carried out by (Broadbent & Poon, 2015), empirical evidence indicates that SRL strategies are significantly related to the final students' academic grades. Out of the 11 studies reviewed where SRL skills such as metacognition, time management, effort regulation, peer learning, elaboration, rehearsal, organization, critical thinking and help seeking were measured, it was found that metacognition, time management, effort regulation and critical thinking are significantly associated with academic achievement on online learning platforms. Broadbent and Poon (2015) suggest that research studies should consider how mediating factors such as motivation and self-efficacy and how they can be measured and modeled from LMS data should be taken into account.

2.2.1 Theoretical Frameworks of SRL

Self-Regulated Learning is founded on various theoretical frameworks that guide research on SRL. In this section, we explore and compare the existing frameworks and describe the different phases, processes, and constructs to provide insights into the nature of SRL and its role in learning and the SRL skills considered in this study. Seven models namely Zimmerman’s (2000) model, Boekaerts’s (2017) model, Winne and Hawin’s (2011) model, Pintrich’s (2000) model, Efklides's (2011), and the Socially Shared Regulated Learning by Hadwin, et. al (2011) model were highlighted. This offered the basis and justification for the choice of the SRL model that was adopted in this study.

2.2.1.1 Zimmerman’s Model

Zimmerman’s 2000 social cognitive model which is commonly referred to as the cyclical model comprises three phases (Núñez et al., 2017);

a. ***The forethought*** phase is where students analyze tasks, set goals, and plan on how to achieve the goals through implementing active learning strategies.
b. ***Performance*** phase where students execute learning tasks and monitor their progress through self-control to remain cognitively engaged and motivated
c. ***Self-reflection*** phase involves students’ assessment of task performance and generating reactions for the future.

Several instruments have been developed to validate this model including the Self-Regulated Learning Interview Schedule (SRLIS), Micro analytic measures and the Academic Self-Regulation Scale (A-SRL) (Núñez et al., 2017).

Figure 2.1 presents Zimmerman's SRL Model.

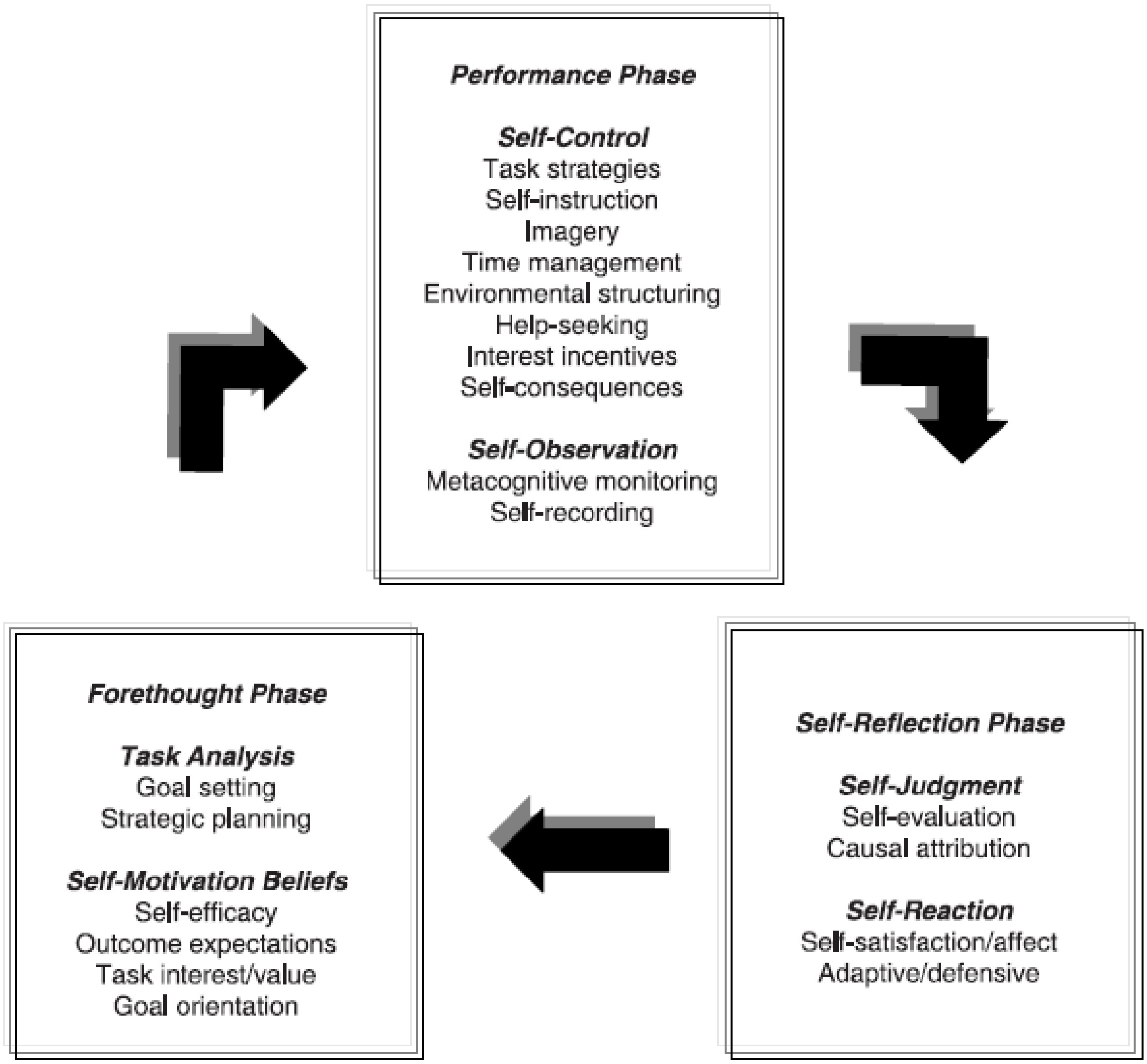


Figure 2.1: Zimmerman's SRL model

2.2.1.2 Boekaerts' SRL Model

Boekaerts' SRL model is also referred to as the dual processing self-regulation model. The model is based on social psychology where learner emotions form the core component of the model. The model describes two goals for learners; first extending students' knowledge to improve their cognitive skills called the *growth pathway* and second involves students striving to maintain their steadfastness in learning called the *well-being pathway*. The model describes how students set themselves to achieve the two goals (Boekaerts, 2017).

Boekaerts' model is guided by three objectives for self-regulation;

a. Expanding knowledge and skills,
b. Preventing threats to the self and loss of resources and
c. Protecting one's commitments.

The studies that have been carried out in support of the model mainly used the On-line Motivation Questionnaire (OMQ), Interactive Learning Group System (ILGS), Confidence and Doubt Scale and Neural Network Methodology as the measurement tools. These tools were mainly developed by Boekaerts and her research teams (Núñez et al., 2017).

Figure 2.2 presents Boekaerts' SRL model.

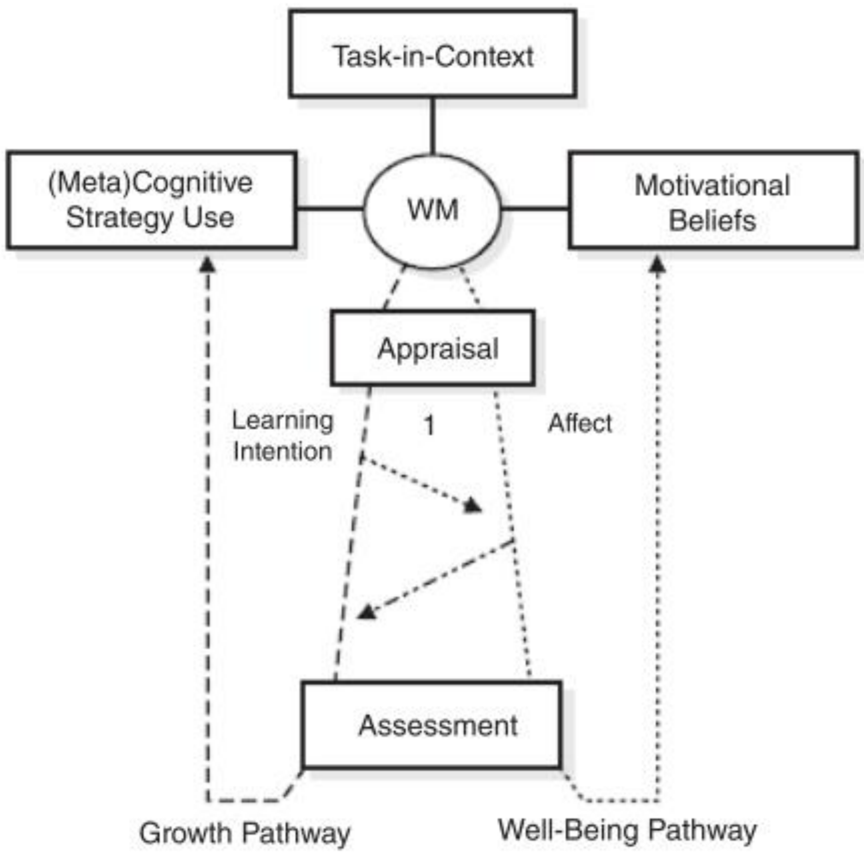


Figure 2.2: Boekaerts' SRL model

2.2.1.3 Winne & Hadwin's Model

Winne & Hadwin's model is grounded on metacognition and is based on the information processing theory. The model comprises four phases and places students as agents who set their own goals in a learning environment and put efforts to achieve the goals (Winne, 2011). The four phases include;

a. ***Task definition*** is where learners try to understand the learning environment and defines the learning tasks.
b. ***The goal setting & planning*** phase involves students identifying academic goals and establishing plans on how to achieve goals.
c. ***The studying tactics*** phase involves students employing strategies to enable them to achieve the goals set in phase two.
d. ***Adaptations to the metacognition*** phase where learners use metacognitive strategies to adapt to any changes to complete the tasks in phase one.

Winnie and Hadwin's model treats learners as agents who choose what information to process and how to operate the learning environment without being affected by the environment. Information processing occurs within each of the phases. The model also describes SRL as progressive as it unfolds during a learning episode and so will allow for measuring SRL using trace logs to measure level of improvement. The model will also help us understand how learners are engage and interact with learning environment and help us tailor learner interventions to advance SRL skills (Núñez et al., 2017).

There are no classical instruments that have been used measure SRL although intervention tools have been developed based on the theoretical framework of the model (Núñez et al., 2017).

Figure 2.3 presents Winne & Hadwin's Model.

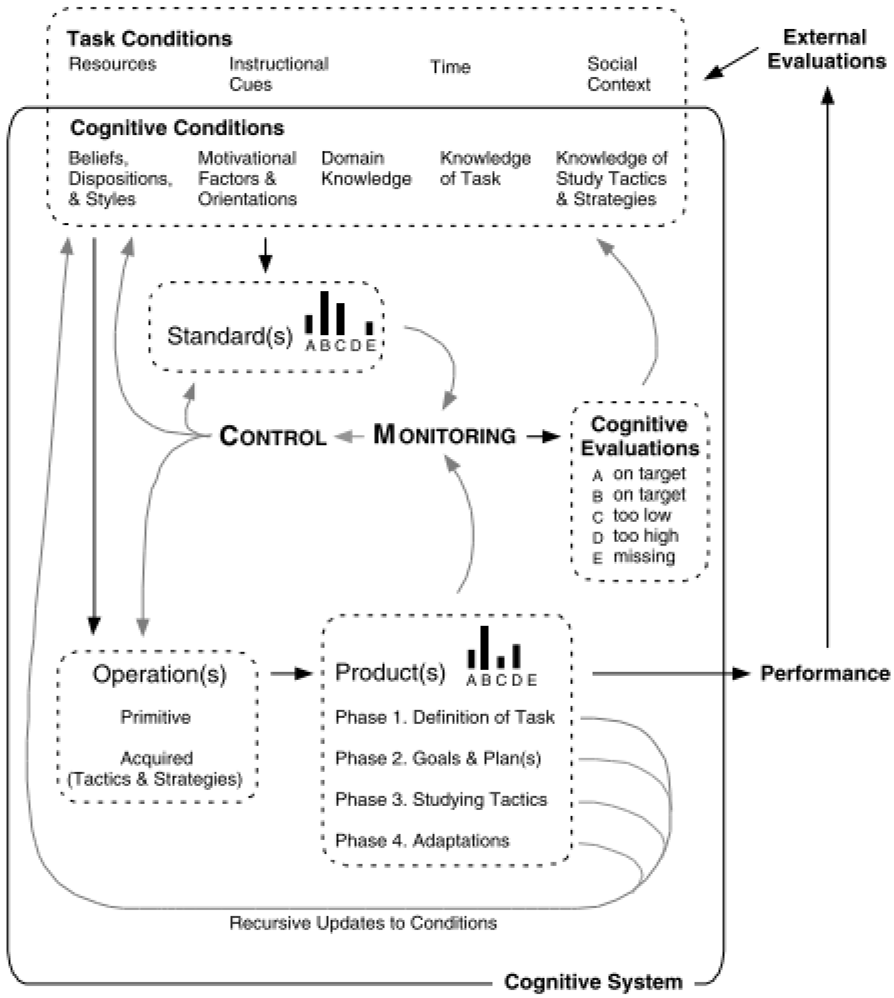


Figure 2.3: Winne & Hadwin's SRL model

2.2.1.4 Pintrich's Model

Pintrich's (2000) SRL model is grounded on social cognition and motivation. The model describes four phases and four areas of self-regulation as presented in Figure 2.4. The phases of regulation include;

a. ***The forethought, planning, and activation*** phase require that students' knowledge and perception of tasks and context are activated.
b. ***The monitoring*** phase involves the monitoring of metacognitive awareness of tasks and the contextual environment.

c. ***The control*** phase is concerned with the control and regulation of different tasks during the actual learning process.

d. ***The reaction and reflection*** phase involves the reactions and reflections that learners will take depending on the outcomes of the learning tasks and context. The four areas that learners are supposed to monitor, control and regulate in each phase are cognition, motivation, behavior, and context where learning occurs. According to (Pintrich & Zusho, 2002), SRL is influenced not only by metacognition, behavior, and motivation but also by the contextual environment where learning takes place.

Each phase describes four areas of self-regulation: cognition, motivation, behavior, and context. The SRL model describes the learner's attempt to control their behavior during the learning process. Pintrich's is the only model that captures learner behavior (Núñez et al., 2017). The MSLQ instrument developed by Pintrich has been used to measure SRL based on Pintrich's model. MSLQ is the most used instrument to measure SRL (Roth et al., 2016b).

Figure 2.4 presents Pintrich's SRL Model.

	Areas for regulation			
Phases	Cognition	Motivation/affect	Behavior	Context
1. Forethought, planning, and activation	Target goal setting	Goal orientation adoption	[Time and effort planning]	[Perceptions of task]
	Prior content knowledge activation	Efficacy judgments	[Planning for self-observations of behavior]	[Perceptions of context]
	Metacognitive knowledge activation	Ease of learning judgements (EOLs); perceptions of task difficulty		
		Task value activation		
		Interest activation		
2. Monitoring	Metacognitive awareness and monitoring of cognition (FOKs, JOLs)	Awareness and monitoring of motivation and affect	Awareness and monitoring of effort, time use, need for help	Monitoring changing task and context conditions
			Self-observation of behavior	
3. Control	Selection and adaptation of cognitive strategies for learning, thinking	Selection and adaptation of strategies for managing motivation and affect	Increase/decrease effort	Change or renegotiate task
			Persist, give up	Change or leave context
			Help-seeking behavior	
4. Reaction and reflection	Cognitive judgments	Affective reactions	Choice behavior	Evaluation of task
	Attributions	Attributions		Evaluation of context

Figure 2.4: Pintrich's SRL model

Pintrich’s SRL model model provides the framework for understanding how learners engage and interact with the learning environment. Table 2.1 describes the reasons for the choice of Pintrich’s SRL model as the conceptual framework to guide this study.

Table 2.1: Reasons for Choosing Pintrich's Model of SRL

Reason	Reference
The model incorporates the learner’s attempt to control their behavior during the learning process	(Núñez et al., 2017) (Pintrich & Zusho, 2002)
The phases are time-ordered and therefore allow learners to interactively engage in more than one phase simultaneously	(Schunk, 2005))
The model takes into account the contextual factors that may influence learning for example the learning environment where learning occurs	(Pintrich & Zusho, 2002)
The model addresses the complexity of SRL while considering the educational settings where learning takes place, especially those taking place outside controlled environments	(Pintrich & Zusho, 2002)
Being a social cognitive model it addresses students’ motivational affordances such as self-efficacy and goal orientation in the context of the learning environment and the influence of peers and instructors	(Whipp & Chiarelli, 2004)
Pintrich’s model is effective in providing interventions that aim at enhancing SRL skills for college and university education	(Núñez et al., 2017)

2.2.1.5 Efklides' MASRL Model

The Metacognitive & Affective Model of Self-Regulated Learning (MASRL) is grounded in social cognitive theory. It is composed of cognition, motivation, self-concept, affect, volition, metacognitive knowledge, and metacognitive skills. The model describes the relationship between metacognition, motivation, and affect through the macro and micro levels interaction (Núñez et al., 2017). Two instruments developed to measure SRL based on MASRL by Efklides and his team; a questionnaire to measure self-concept and the Metacognitive Experiences Questionnaire (MEQ). Figure 2.5 presents Efklides' MASRL model.

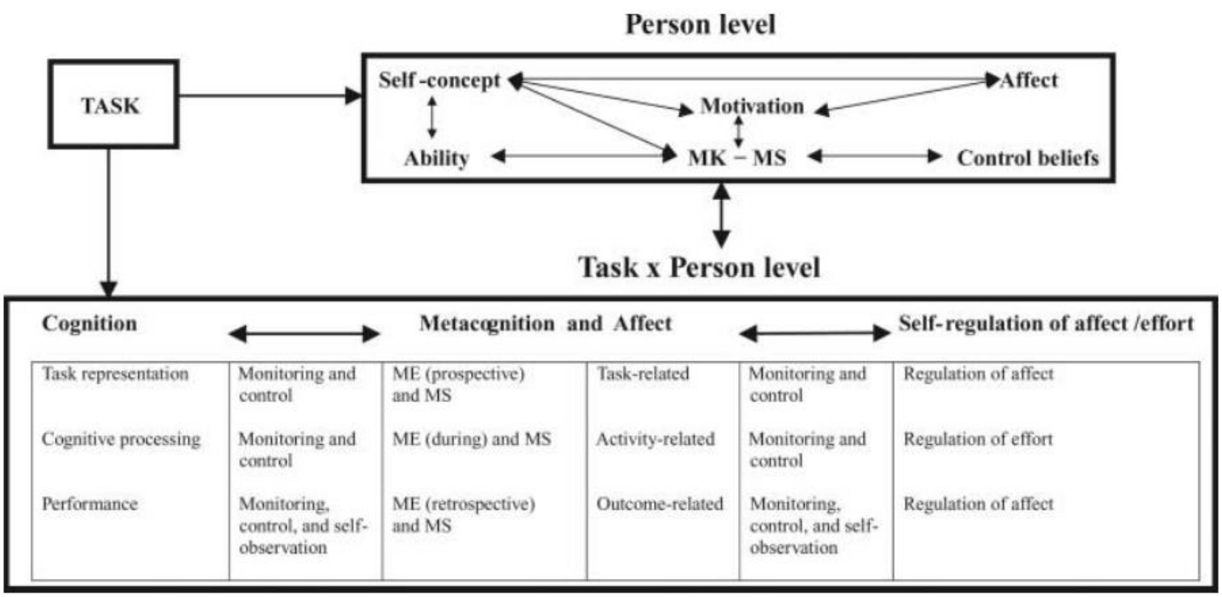


Figure 2.5: Efklides' MASRL model

2.2.1.6 The Socially Shared Regulated Learning (SSRL) Model

Socially Shared Regulated Learning (SSRL) describes SRL in the context of social and collaborative learning. The model explains how to have a successful individual collaboration and hence collective or group regulation by group members where individual regulation contributes to group regulation during learning processes. The model describes three modes of regulation in collaborative environments; self-regulation, co-regulation, and shared regulation (Järvelä & Hadwin, 2013).

Currently there are no instruments have been developed to measure SRL using the SSRL model. Studies use self-reported data in validating this model (Núñez et al., 2017).

Figure 2.6 presents the Socially Shared Regulated Learning model.

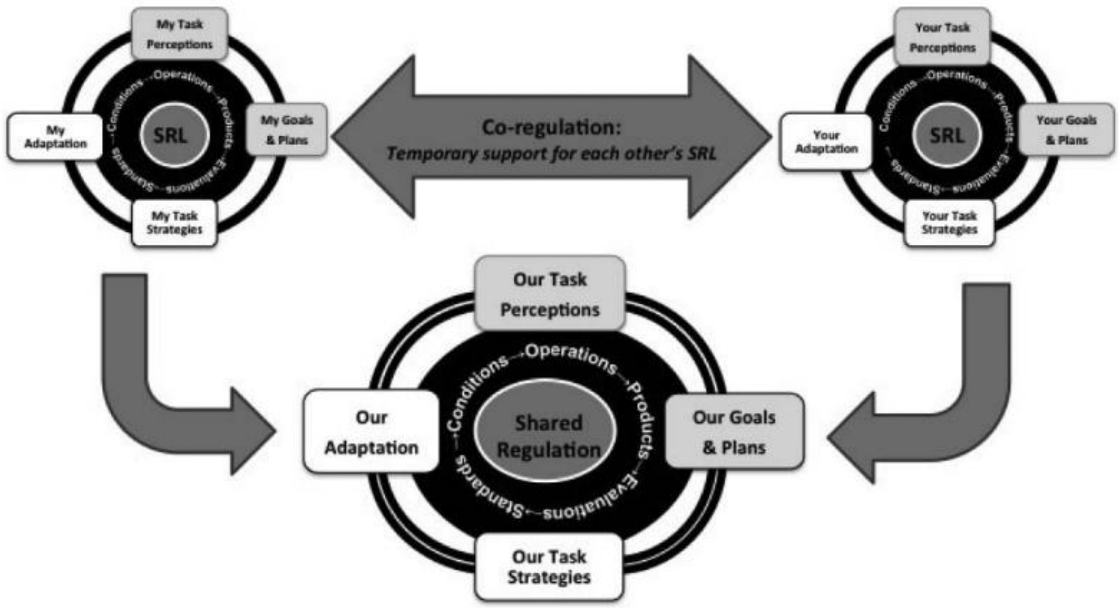


Figure 2.6: The Socially Shared Regulated Learning model

Table 2.2 presents the summary of the SRL theoretical models reviewed.

Table 2.2: Summary of the SRL Theoretical Models

Model name	Associated SRL Component	Model Constructs/Phases
Zimmerman's (2000) Model	Social cognition	• Forethought • Performance • Self-Reflection
Boekaerts' Model (2017)	Social psychology	This model is guided by social psychology where learner emotions form the central part of the model
Winne & Hadwin's Model (2011)	Metacognition	• Cognition • Metacognition
Pintrich's Model (2000)	Motivation	• Forethought, planning, and activation; • Monitoring; • Control; and • Reaction and reflection.
Efklides' (2011) Model	Social cognition	• Cognition • Motivation • Self-concept • Affect • Volition • Metacognition in the form of metacognitive knowledge • Metacognition in the form of metacognitive skills
Socially Shared Regulated Learning (SSRL) Model by Hadwin, Järvelä, and Miller (2011)	Social cognition	The model describes SRL in the context of social and interactive learning through the use of ICT known as computer-supported collaborative learning (CSCL).

From literature, it is evident that there is no single instrument that can be used to measure the SRL process and its constructs as a result of the different SRL models that exist. Recent studies recommend combining models to effectively be able to measure how students employ SRL and how the strategies change over time during the learning process. This proposal may not be easily achieved as this would mean that a complex study is carried out to measure all constructs that will be produced after models have been combined. Secondly, the complexity will be extended since each model is grounded on different aspects of learning (Núñez et al., 2017).

2.2.2 Conceptual Framework for Measuring Self-Regulated Learning Skills

This section describes the conceptual framework used to measure SRL skills in online learning environments and therefore guides the description of the indicators for the SRL strategies that were consider in this study.

The four phases described by Pintrich (2004) provide a foundation for researchers to organize their thinking and concepts about self-regulation. For example monitoring, control, and reaction phases may take place at the same time as students engage in a given learning task. While phase two describes how students maintain metacognitive awareness of self and tasks being learned and phase three explains how students put up efforts to control the self and learning tasks in the learning context and therefore likely to play a role in their study. Moreover, Pintrich (2004) argues that not all academic learning follows the phases of learning linearly. According to Pintrich and Zusho (2002), the phases of SRL in existing models do not affect how learning occurs but provide a heuristic way for researchers to organize their conceptualizing self-regulation in learning processes. This implies that phases do not take place sequentially and therefore can take place simultaneously.

Literature reveals that phases two and three demonstrate developmental outcomes that a learner possesses after attending education in learning institutions and demonstrate that (a) SRL monitoring, control, and reaction may take place simultaneously as a learner engages in learning tasks and (b) that phases two and three are difficult to be empirically separated in an attempt to make clear conceptual distinctions. Therefore, there is a need for studies to focus on the SRL areas that students attempt to monitor, control, and regulate during learning processes. These areas include the cognitive and behavioral strategies of SRL. These two areas can either be influenced by individual learners or external agents such as teachers and peer learners through the instructional direction of scaffolding on how one can handle learning tasks (Pintrich & Zusho, 2002).

2.2.2.1 Regulating Cognitive Strategies of SRL

Cognition is an internal characteristic that takes place within a learner. Cognitive energy engagement can be inferred as the outward behavior of mental engagement during learning ((Halverson & Graham, 2019). According to the conceptual framework by Halverson & Graham (2019), cognitive strategies can be classified into two; quantitative and qualitative strategies. These cognitive strategies reveal the mental efforts put towards learning by the learner. Quantitative strategies are those that are visible and can be manifested outwardly to reflect the mental engagements during learning. They include attention, effort regulation, and time on task as explained below;

- **Attention** refers to the perceptual and processing energies that learners engage during a learning process. Traditionally in the face-to-face classroom, attention is measured through observation while tracking eye movements and body language are some of the methods used to measure attention in online or blended learning environments.
- **Effort & persistence** refers to the engagement level of the learner when faced with learning difficulty for instance revisiting or spending extra time on learning tasks, paying attention to hints, and completing learning tasks on time. Effort and persistence can be measured in terms of time spent when a learner engages in difficulty learning material again and again.
- **Time on task** refers to the amount of time a learner spends handling or solving given learning activities for example time viewing content. Time on task can be measured in terms of the time spent on content pages and resources.

Qualitative strategies include cognitive and metacognitive strategies that keep concentration and individual interest in learning. The qualitative strategies are manifested when learners actively engage their thoughts in their learning. Since literature indicates that metacognitive strategies are linked with behavioral strategies, the indicators of quantitative strategies are most appropriate for online learning.

Self-regulated learners should monitor and control their cognition by being aware of the learning progress and be able to regulate themselves especially when they encounter underlying environmental challenges, hence maintaining an upward trajectory. According to Pintrich (2004) monitoring and controlling cognition help the learner to consciously be aware of what they are supposed to achieve and determine where they are currently. The learner then makes attempts to control and regulate their learning activities to help them to adjust to achieve the goals set. These learning activities for regulating cognition include rehearsal, elaboration, organization, and critical thinking as explained below;

- **Elaboration** is the ability of students to be able to relate new knowledge to existing knowledge while engaging in learning activities such as interpretation and summarization.
- **Rehearsal** refers to the ability of the learner to learn by going through learning material repetitively to be able to remember it when necessary. Rehearsal can also be expressed when learners view learning content again and again during the online learning process.
- **Organization** is the ability to select relevant and appropriate learning material and be able to link the information provided by the material.
- **Critical thinking** is the ability to relate existing knowledge to new situations during the learning process.

2.2.2.2 Regulating Behavioral SRL strategies

In self-regulated learning, regulation of behavior involves individual learners making efforts to control their learning habits. The SRL strategies that relate to learning behavior include time management, help-seeking, and effort regulation.

- **Time management** involves the learner scheduling time to engage in academic activities such as studying. For example, students taking the initiative to manage their study schedules and learning resources within the learning environment;
- **Effort regulation** involves students controlling the amount of energy put towards achieving course goals;
- **Help-seeking** describes those students who can know when to seek assistance from students or instructors within a given contextual learning environment especially when confronted with difficulties in learning. Help-seeking may also refer to the social interaction within a learning environment with other peers.

2.3 Theories of Learning

Learning theories exist to help us conceptualize how learning and teaching take place and also explain how knowledge is acquired by students. This section discusses three main theories of learning namely behaviorism, constructivism, and cognitivism, and highlights the theory that guided this study.

2.3.1 Behaviorist Theory of Learning

The behaviorist theory of learning proposed by Skinner (1938) is grounded on the concept that learning causes a change of behavior in the learner when experience is gained from the surrounding environment. Behaviorists, therefore, view learning as a passive absorption of knowledge as a result of a response to stimuli from an external environment (Araiba, 2020). According to this theory, the teacher is at the center of the learning process and learning is viewed as an independent process that takes place outside the learner which starts to take place when learners are provided with experience. Learners' behavior changes due to a response from a stimulus that can be acquired habits. Diverse techniques such as reinforcement, consequences, and contracts are used to reward students for desirable behaviors and punish them for undesirable behaviors. When these techniques are applied to learners, their behaviors can change and bring them satisfaction, positive feelings, and success in their academics (Zhou & Brown, 2015).

2.3.2 Cognitivism Theory of Learning

The cognitivism theory of learning describes learning as the active processing of information and acquiring knowledge by the learner as contrasting to the change of behavior (Mandler, 2002). The cognitive theory describes how learners are taught to observe a variety of learning skills through thinking. The information that surrounds the learners is processed into knowledge and stored in the memory (Grider, 1993). According to Sweller (2011), when students are left to learn on their own, a high cognitive load may be experienced. However, to counter this problem, students need to be provided with learning interventions that enable them to avoid overloading their cognitive processes. Self-regulated learning relates to the cognitive theory of learning since it allows learners to assume active roles in learning processes while being guided by their motivation and metacognitive skills. Since the cognitive theory gives learners the freedom to engage in their learning, this study adapted the cognitivist theory of learning both in the design and development EDM model and the conceptualization and delivery of the EDM-based interventions that were utilized by students in improving cognitive skills. The interventions (that were delivered inform of prompts and visualized feedback) considered students' abilities and afforded them with opportunities that enabled them manage their own learning processes in terms of time management, self-monitoring and help-seeking (Sweller, 2011).

2.3.3 Constructivist Theory

The constructivist theory which was founded by Piaget (1972) is grounded on the belief that learners can obtain and construct knowledge based on their previous experience (von Glasersfeld, 2014). According to constructivism theory, learners construct their knowledge at both individual and collective levels guided by the experiences they receive and the reflections upon those experiences. Since constructivism theory views the learner as an active participant in the learning process, the environment within which knowledge is constructed should be considered as the one from which students receive experiences (Xu et al., 2014). The constructivism theory assumes that the learner is at the center of the learning process and this might help them to become SRL learners (Amineh & Asl, 2015; Bada, 2015; Huang, 2002). The learner characteristics from a constructivist perspective fit into strategies of self-regulation in learning where learners utilize their prior experiences and knowledge to make meaning out of their learning. From the perspective of SRL, learners need to be aware of their learning and take charge of their learning process. In this study therefore we seek to describe how the principles of constructivism can be articulated in SRL within LMS. Although a majority of LMS systems were designed taking into consideration, knowledge constructions, still learners are treated as passive recipients of information, and learning activities are those that support one-size-fits-all (Xu et al., 2014). A self-regulated learner can be able to stay active throughout the learning process.

E-learning environments should be designed and reinforced to provide learners with opportunities to learn actively as they create knowledge as this is the principle advocated by both constructivism theory of learning and SRL (Bada, 2015). Although literature indicates that efforts are going to create learning environments that reinforce the constructivism theory of learning, not all the principles have been addressed; For example, Maina et al. (2017) designed an intelligent tool to support group formation and enhance the online collaborative learning which is one of the characteristics that a constructivist learning environment should possess (Tam, 2000). Literature indicates that current learning environments are yet to achieve the ideal characteristics of constructivist learning environments as envisaged by the worldview of constructivists. Concerning this study therefore the cognitivist and constructivist theories of learning were adopted. The summary of the ideal characteristics and the current challenges identified in literature are highlighted in Table 2.3.

Table 2.3: Shortfalls of the current Constructivist Learning Environments

Ideal Characteristic	What literature indicate
Knowledge should be shared between teachers and students	Students are active recipients and processors of information from teachers (Bogarín et al., 2018b; Luna, Castro, & Romero, 2017)
Teachers and students share authority	Learning is teacher-centered. Many institutions of learning utilize LMS to deliver learning materials and information rather than enabling learners to actively engage in learning processes (Bogarín et al., 2018a; Luna et al., 2017).
The teacher’s role is one of a facilitator or guide	Automation is yet to be achieved to allow teachers to guide. Research going on allows teachers to guide or support learners based on analyzed data from educational environments (Pardo et al., 2018).
Learning groups will consist of small numbers of heterogeneous students	Maina et al. (2017) designed a tool based on machine learning to support group formation based on various factors such as learner competencies.

2.4 Measurement of Self-Regulated Learning Strategies

Since the emergence of SRL in the 1990s, various tools and techniques have been developed and used to measure self-regulatory skills for both online and face-to-face teaching and learning. These methods are discussed in this section while taking into consideration the historical progress, the measurement tools that have been used in the past, and how helpful they have been concerning measuring SRL presently and give direction where research especially on measuring SRL for e-learning environment needs to go.

2.4.1 Trends in Measurement of Self-Regulated Learning

The measurement of SRL strategies is carried out to establish students' levels of self-regulatory skills or to what extent learners self-regulate their learning (Triquet et al., 2017). The measurement process is conducted to establish the extent to which students utilize SRL skills such as help-seeking, peer-to-peer learning, time management, effort regulation, metacognition, elaboration, organization, self-monitoring, self-evaluation, and critical thinking in terms of comprehension and progress awareness (Araka et al., 2020; Broadbent & Poon, 2015) using established tools and instruments.

The SRL measurement tools can be categorized into two namely offline and online tools. The offline tools measure SRL before or after a learning episode through the use of self-report tools. On the other hand, online tools are used to measure SRL skills during the learning episode. An example of an online tool or technique is the use of trace data collected from a learning environment. Compared to offline tools, online tools are viewed as unobtrusive since the tools are deployed without learners being aware and hence online tools do not affect learners' engagement behavior and performance (Schraw, 2010). Offline tools which are commonly referred to as self-report measures include such tools as Online Self-Regulated Learning Questionnaire (OSLQ), Motivated Strategies for Learning Questionnaire (MSLQ), Learning and Study Strategies Inventory (LASSI), Self-Regulation Measure for Computer-based Learning (SRMC) and Self-Regulated Learning Interview Schedule (SRLIS). These tools have been in use since the emergence of the SRL theory. At its inception, SRL was viewed as students' character traits and did not consider the environment where learning took place. As a result, self-report tools such as structured questionnaires and interviews were developed to measure SRL before or after a learning process. Winne and Perry (2000) observes that self-report tools face major drawbacks; the predisposition for students to overrate their SRL skills hence the outcome is dependent on learners' perception of their SRL skills. Also, the offline tools are deployed outside a learning environment (Roth et al., 2016a; Winne & Perry, 2000). Lee (2008) opines that when measuring SRL, there is a need to take into consideration the learning environment within which learning occurs. The use of self-report tools does not allow for the provision of interventions for learners. Thus, little knowledge is known about how to provide SRL interventions based on SRL evaluation results based on the offline tools.

The introduction of online learning as a continuum of traditional face-to-face teaching and learning changed the concept of viewing SRL as trait-based to a process-oriented task to be measured as a series of events. Online tools include the use of log traces collected from an online learning environment. Most studies rely on the use of data generated from classroom observations, laboratory studies, self-test scores, self-reporting techniques, and surveys to measure the level at which learners employ SRL in online learning environments (Kizilcec et al., 2017; Saks & Leijen, 2014).

Despite the challenges raised by many studies, self-report tools have continuously been used to measure SRL in both traditional classroom setups and online learning environments. The self-report tools, which were originally developed for use in face-to-face learning set-up, are used by students in recording their perceived SRL traits and their utilization in the learning process. The continued use of self-report tools may be due to their validity and reliability that has been tested and improved over time (Roth et al., 2016a). The self-report tools, however, may not apply in measuring SRL in online learning environments. The summary of these challenges is presented in Table 2.4.

Currently, we are witnessing the development and deployment of SRL tools not only measure SRL but also provide interventions for supporting/promoting SRL skills in learners. This is being made possible as a result of the emergence EDM (Panadero et al., 2016). The EDM approach is seen as an unobtrusive technique that can be used to provide insights into how students self-regulate in online learning environments compared to the use of self-report tools. Additionally, researchers are also recommending that there is a need to measure SRL as learning progresses and this will be achieved through the analysis of learner data collected from educational environments such as LMSs (Winne & Baker, 2013).

Table 2.4: Challenges Faced when Measuring SRL in Online Learning Environments

Challenge	Description	Reference
Use of traditional tools/methods in e-learning	The traditional instruments such as questionnaires and interviews are trait-based and user-oriented; learners respond to SRL items depending on how they perceive themselves leading to learners overestimating their use of SRL skills The tools are also deployed outside the learning environment before or after a learning episode and therefore not able to measure SRL during an actual learning episode when skills are being employed by students.	(Broadbent & Poon, 2015; Lee, 2008; Roth et al., 2016a; Saks & Leijen, 2014; Siadaty, 2016; Winne & Perry, 2000)
The traditional tools and methods are obtrusive	The learners are normally aware of SRL being measured and therefore affects their engagement and performance	(Schraw, 2010) (Siadaty, 2016)
Existence of many models and many constructs to be measured	There is no generalized model that describes or conceptualizes all SRL constructs. Additionally, each of the existing models is grounded on different aspects of learning.	(Núñez et al., 2017)
Lack of a model for both SRL measurements and promotion	What next after establishing one's level of SRL? So far we have had separate tools for measuring and promoting SRL.	(Panadero et al., 2016)
Lack of framework that describes/guides how to establish learners' levels of SRL and also that which describes at what level to start and stop issuing scaffold	The existing theoretical models only provide frameworks that describe the different phases, processes, and constructs to be measured. When it comes to actual measurement and provision of scaffolds there is no defined framework to follow for guidelines.	(Araka et al., 2020)

2.4.2 Students' SRL profiles in Online Learning Environments

Student engagement behavior in online learning environments is being used as an indicator of how students employ self-regulatory skills during learning processes. Levels of online engagement as inferred from students' log data activities can therefore be used as indicators for the various levels of SRL. For instance, an active student can be identified through his participation in academic activities and visits to learning resources. The implied observation is the student is in control of his learning process and therefore exhibits a high level of self-regulatory behaviors.

Students' engagement behaviors and learning patterns in online learning environments such as LMS can be measured using trace data. The dataset features may include content or page views, frequency of logins, access to learning materials, forum posts by students, and quiz and assignment scores (Araka et al., 2020). Previous studies indicate that distinct profiles of SRL exist among students who engage in online and blended learning. The profiles can be identified using educational data mining methods applied to self-report data, trace data, or both (Barnard et al., 2010). For example, latent class analysis was used to identify five different profiles in SRL. In this study, a self-report online questionnaire known as Online Self-Regulated Learning Questionnaire (OLSQ) was used to collect self-report data which was used to identify five profiles of self-regulators; super self-regulators, competent self-regulators, forethought-endorsing self-regulators, performance/reflection self-regulators and non/minimal self-regulators (Barnard et al., 2010).

In a related study, trace data comprising logs related to access to learning materials, completion of quizzes, and answer logs was analyzed for profiles in SRL. From the data, various behaviors were measured including the number of completed quizzes, total access time, reviewing time, scores of completed quizzes, anti-procrastination, and irregularity of study interval and pacing (Li et al., 2018). The K-means clustering algorithm was applied to the data and four distinct clusters were identified; early completers, late completers, early dropouts, and late dropouts. However, the data only comprised assessment data which did not indicate the student interactions with the course. The students' activities were only limited to listening and reading and this may not reflect the actual learner behaviors in an actual online learning environment.

A mixed approach where both trace data and self-report data were used to profile online learners into three clusters: high self-regulators, medium self-regulators, and low-self regulators. While trace data was used during the analysis, the self-regulated learning skills that were analyzed and used to classify learners were reported by the learners in various stages during the study period. The two-step cluster analysis was used to group the learners: the first step was the pre-cluster formation and the second; the hierarchical clustering algorithm was used to merge the pre-clusters leading to the three distinct groups (Ainscough et al., 2019). The trace data used in the study comprised of average word count for each meta-learning question, submission time for the meta-learning tasks, and completion rate of the tasks.

Finally, Çebi and Güyer (2020) presented various learning activities to learners using Moodle LMS. The learning activities include a tutorial, video, concept maps, exercises, summary, highlight, and forum activities. The dataset used was collected from three sources: MSLQ, trace data, and assessment data. Cluster analysis involving hierarchical clustering and K-means were used to identify three clusters. The study however was limited to only three weeks of study on a single course and therefore may not have given proper observance of behavior change in learners as far as SRL is concerned. Table 2.5 presents the summary of the SRL profiles identified from online learning using self-data, trace data, or both and the techniques used to identify the SRL strategies.

Table 2.5: Summary of SRL profiles identified from Literature

Reference	SRL profile	Data Source	The technique used to identify the profiles
(Valle et al., 2008)	• Intermediate SRL level • High SRL level • Low SRL level	Self-Report	Two-step cluster analysis
(Barnard et al., 2010)	• Super self-regulators, • Competent self-regulators, • Forethought-endorsing self-regulators, • Performance/reflection self-regulators and • Non/minimal self-regulators	Self-Report	Latent class analysis
(Yot-Domínguez & Marcelo, 2017)	• High-level regulators • Low-level regulators	Self-Report	Stepwise cluster analysis • Hierarchy analysis • Ward method • K-means analysis

(Gašević et al., 2017)	• Formative assessment, • Summative assessment through trial and error • Studying reading materials • Video watching with formative assessment	Self-Report and Trace data	Agglomerative hierarchical clustering (based on Ward's algorithm)
(Li et al., 2018)	• Early completers • Late completers • Early dropouts • Late dropouts	Trace data	K-means clustering algorithm
(Broadbent & Fuller-Tyszkiewicz, 2018)	• Minimal regulators • Restrained regulators • Calm self-reliant capable regulators • Anxious capable collaborators • Super regulators	Self-report	Latent profile analysis
(Kim et al., 2018)	• Self-regulation • Partial self-regulation • Non-self-regulation	Trace data and Self-report	K-medoids clustering
(Ainscough et al., 2019)	• High self-regulators • Medium self-regulators • Low self-regulators	Self-report and Trace data	Two-step cluster analysis
(Peach et al., 2019)	• Early Birds • On-time • Low Engagers • Crammers • Sporadic outliers (unclustered learners)	Trace data	Mathematical framework (based on dynamic time warping kernel & clustering algorithm
(Çebi & Güyer, 2020)	• Cluster 1: Students with the least interaction • Cluster 2: Intense interaction with video, example, and forum activities • Cluster 3: Students who spend more time on tutorials, exercises, concept maps, summaries, and highlight activities	Self-report and Trace data and assessment data	Cluster analysis • Hierarchical clustering • K-means clustering

2.5 Promoting Self-Regulated Learning in Online Learning

Self-Regulated learning interventions refer to activities that can "trigger SRL development" within an online student during a learning episode to enhance students' SRL skills (Triquet et al., 2017). The interventions stimulate the growth of SRL skills for learners hence beneficial in enhancing learner involvement. Providing SRL interventions helps students to plan, monitor, and reflect on their learning habits by providing relevant guidelines and cues to improve their self-regulatory skills, especially in online learning environments (Viberg et al., 2020).

The measurement and promotion of SRL can be described as having undergone three stages. The first era is when different methods and instruments were combined to assist infer SRL strategies. The second era was the use of online measures characterized by the use of online data within the context of learning environments. In early work on the subject, SRL was conceptualized as an individual inclination that was based on learner traits without contextual considerations and this led to the development of trait-based tools. Later this concept improved when researchers discovered that students' SRL strategies were dependent on the domain or learning environment where the learning occurred. Trait-based measures were therefore replaced with domain-based measures currently referred to as online measures where SRL is measured during the learning process (Boekaerts & Cascallar, 2006). To this level, we have witnessed measurement tools that are separate from intervention tools. The third era is the proposed one where we are now. In this era, measurements and interventions are supposed to be designed so that they can co-occur. Recent studies propose SRL measurement tools that also promote SRL at the same time (Lodge et al., 2019; Panadero et al., 2016).

Recent studies recommend the nature and type of interventions that students of higher education especially e-learning students should receive in relation to the type of models that deliver effective interventions. According to (Núñez et al., 2017), each of the different categories of learners from the Primary level to those of higher education is supposed to receive interventions that impact their learning and should be based on certain types of SRL theoretical framework. SRL interventions have different effects/usefulness depending on the education level of the learner/recipient of the intervention. Interventions that promote learning besides assisting the learners to develop SRL skills. There is also a need to provide interventions that reinforce group work as grounded on the Socially Shared Regulated Learning (SSRL) model to encourage collective self-regulation among students and encourage the social construction of knowledge which is important in improving learner motivation. The SRL theoretical models work well for different levels of learners ranging from primary school pupils to higher education students (Núñez et al., 2017).

Interventions that are grounded on Social Cognitive theory have been found to have a greater impact on primary school pupils. Secondary school students benefit more from those interventions that are based on metacognition. The reason for this could be because of the growth of cognitive skills and for this, SRL models that are based on metacognition such as Efklides's (2011) and Winne & Hadwin's (2011) model are likely to work well for this education level. For the higher education and professional workers level the use of social cognitive models is likely to have a higher effect. (Núñez et al., 2017) argue that this class of learners needs multifaceted interventions that aim at improving their motivation and emotional aspects of learning. The interventions for this level should be provided through models like Boekaerts (2017), Pintrich (2000), and Zimmermann (2000) models (Núñez et al., 2017).

Table 2.6 presents the summary of the effectiveness of SRL interventions on learners' education levels.

Table 2.6: Effectiveness of SRL Interventions on Learners' Education Levels

Level of Education	SRL theoretical framework	Type of interventions
Primary	Zimmerman's (2000) Model	Interventions grounded on Social cognitive theory
Secondary	Efklides (2011), Winne Hadwin (2011) models	Interventions based on metacognition
Higher education	Boekaerts (2017), Pintrich (2000), and Zimmermann (2000) models	multifaceted interventions that aim at improving their motivation and emotional aspects of learning

The underlying indicators of providing interventions that can lead to the provision of real-time feedback on students' learning behaviors using Educational Data Mining (EDM) have not been addressed. The studies have also not addressed how EDM can be used to address how interventions can be used to promote SRL consistently and hence lead to mastery and retention (Iwamoto et al., 2017). According to Viberg et al. (2020), there is limited empirical evidence of LA/EDM being used to provide support for SRL. The existing studies analyze data obtained from learning environments to establish students' learning behaviors rather than supporting SRL among learners.

Literature indicates that there is research going on towards the provision of SRL interventions based on EDM and LA (Liu et al., 2017; Pardo et al., 2018). However, there is no model for reference that takes into consideration the measurements together with interventions, group learner activities, the SRL model for the best group, and interventions geared towards learning.

2.5.1 Techniques of Promoting SRL in Online Learning Environments

In this section, we discuss the various techniques that are used to promote SRL in online learning environments: prompts, feedback, integrated support systems, and training of learners (Devolder et al., 2012; Rowe & Rafferty, 2013; Wong et al., 2018). These techniques are discussed and the rationale for their use explained.

a. Prompts

Prompts are provided to learners in online learning environments such as LMS and MOOCs in form of questions, advisory messages, instructions, or suggestions to guide and support the students in performing learning activities geared towards employing SRL strategies in learning. The prompts are meant to sustain learners and help them to reflect on and continue employing SRL strategies before, during, or after a learning process (Lehmann et al., 2014; Wong et al., 2018). These prompts, which are more effective in promoting personalized and adaptive learning, can be categorized into two; generic and directed prompts.

a. Generic prompts encourage students to willingly stop their learning and reflect on the learning processes. Generic prompts, therefore, give learners the freedom in exercising autonomy during learning.
b. Directed prompts on the other hand "compel" learners to stop learning and reflect on their learning processes. The directed prompts are more effective, especially for novice learners who have little knowledge about self-regulation.

Prompts are the most effective technique for promoting SRL in online learning environments such as LMS and MOOCs (Devolder et al., 2012; Lehmann et al., 2014; Wong et al., 2018). Prompts in form of questions or recommendations such as learning hints allow students to stop learning and reflect on their thoughts regarding behavior and efficiency of the strategies they are employing by guiding them on specific learning activities (Rowe & Rafferty, 2013; Wong et al., 2018). In supporting cognition, prompts have mostly been used. Various forms of prompts such as guiding questions, advisory promptings, look-back prompts, and reflective prompts are used. In supporting behavior, prompts in form of guiding questions have been used (Devolder et al., 2012; Wong et al., 2018).

Table 2.7 presents the list of the prompt messages that were identified and were used to promote SRL in Moodle LMS in this study.

Table 2.7: List of Prompts used to Promote SRL

SRL Strategy	SRL Strategy	Corresponding SRL Intervention Messages	Reference
Cognition	Time of Task	• Allocate and spend adequate time for each learning activity including chats, discussion forums, and quizzes • Did you know that there are some learning activities that you have not accessed or participated in and the deadline is passing? • Time yourself and have enough time for each learning activity • Did you know that putting more effort and time into difficult areas of study helps you understand the course? Avoid decreasing your effort as you go along.	(Schumacher & Ifenthaler, 2020) (Wong et al., 2020) (van Merriënboer & de Bruin, 2019)
	Rehearsal	• If you are facing any difficulty understanding this course material, consider re-reading the difficulty part • Remember to pause and remind yourself what you are studying, and if you are understanding the course module • After you read a module try to summarize what you have read • Did you know that you can pause reading and think about previous content already covered • Try to relate what you have covered in this lesson with what you already know	(van Alten et al., 2020) (Schumacher & Ifenthaler, 2020) (Wong et al., 2020)

Behavior	Time Management	• Allocate time to go through the course content/module • Allocate time and go through the discussion forums posted by the course Instructor • Allocate time and go through the chats posted by the course Instructor • Allocate time and attempt a scheduled quiz	(Schumacher & Ifenthaler, 2020)
		• Try to allocate time to study this course and follow the schedule of learning activities • Are you following the given schedule? Try complete learning activities on time. • Are you spending enough time on each learning activity?	(Wong et al., 2020)
	Effort Regulation	• Think through as you respond to questions and make notes	(van Alten et al., 2020)
		• Consider pausing and reflecting on the content you are reading and make your notes	(Schumacher & Ifenthaler, 2020)
		• As you go through this content, try taking time to reflect and recall what you are studying • At the end of each lesson, try to review what you have already covered • Remember to spend enough time reviewing course content and participate in learning activities • Try to put more effort to study course resources and participate in learning activities posted by the instructor • Are there areas or learning materials that are difficult to understand? Try and revisit them.	(Wong et al., 2020)

Help-seeking	• Consider seeking assistance from peers or instructors when you experience difficulty in learning • Allocate time to view and respond to forum posts created by course your peers • Allocate time to view and respond to chat posts created by course your peers • Try to allocate specific time to respond to other students' forum posts and chats • Are there areas or learning materials that are difficult to understand? Try and seek help from the course instructor or other students through messages, chats, or forums	(van Alten et al., 2020) (Schumacher & Ifenthaler, 2020) (Wong et al., 2020)

b. Feedback

Feedback is an SRL promotion technique that makes learners aware of their learning processes to take appropriate measures in improving their learning. The feedback can be presented to learners inform of visualizations such as graphs presented to learners through dashboards. Feedback when implemented alone is a less effective approach compared to prompts hence researchers argue that there is a greater effect of combining feedback and prompts (Wong et al., 2018). One of the reasons could be because feedback is passive and students may not be aware of the reason why the feedback is presented to them.

c. Integrated software agents

Integrated software agents are support systems that are embedded with features that promote various SRL strategies through the use of prompts, feedback, or a combination of both. These software agents are built into online learning platforms such as web-based systems and Intelligent Tutoring Systems (ITS) and are meant to scaffold learners' SRL skills. According to Wong et al. (2018), it is only when students use such integrated software agents that their SRL skills will be enhanced. This implies that when students are not oriented or trained about the integrated software system features, they do not utilize them, and therefore may not be effective in implementation.

d. Training of Students

Training of students is another method through which SRL can be promoted amongst the learners. In training, SRL strategies such as cognition, metacognition, and motivation are instructed to learners. Instructors facilitate training by providing learning materials to learners to help them understand the existence and importance of SRL strategies and the need to apply them in their learning. Formal training can also be organized by schools and departments of universities (Wandler & Imbriale, 2017; Wong et al., 2018). With this approach, students are trained and allowed to start utilizing the skills they have been trained on (Rowe & Rafferty, 2013).

2.6 Learning Management Systems

Learning Management Systems (LMS) have been widely used by institutions of higher learning to facilitate teaching and learning (Şahin & Yurdugül, 2020). As a result, a huge amount of log data is generated which if analyzed could be used to inform how learners behave online and help to provide scaffolds to enhance their learning process. Active SRL learners can engage with peers, spend time online learning, and can utilize features LMS features to learn and communicate with peers and instructors (Lee et al., 2019).

Institutions of higher learning that use Learning Management Systems (LMS) such as Moodle and Blackboard in teaching utilize the LMS platforms to deliver course materials and information rather than enabling learners to actively engage in learning processes (Bogarín et al., 2018b; Luna et al., 2017; Romero et al., 2008). Additionally, the LMSs do not provide tools that can enable tracking and evaluating learner activities to evaluate the learning process (Dabbagh & Kitsantas, 2013; Dutt et al., 2017).

Learning Management systems contain inbuilt features and functionalities such as discussion forums, chats, quizzes, assignments, grading, emails, and wikis which are intended to make learners active participants in learning. However, the features are underutilized by learners, and e-learning environments are only used as delivery channels for content to students regardless of their underlying behaviors and characteristics and are therefore not likely to promote the development of SRL. Therefore, students need to employ much more effort which requires students who can self-regulate. Although the utilization of these features does not directly relate to SRL, researchers argue that if well utilized by students during learning, they stimulate the growth and development of SRL skills (Vovides et al., 2007). Although LMS has a rich set of features and functionality that are supposed to support learners in engaging with learning activities by capturing learning activities into databases, the log data is stored in low-level formats that cannot be useful without analysis (Luna et al., 2017; Romero et al., 2008). This is also coupled with a lack of mechanisms on how to support and guide students in enhancing learning and teaching in relation to the existing learning theories such as SRL. There is also a lack of empirical evidence to establish the effect of interventions on learners on utilizing the LMS features and their effect on academic achievement (Araka et al., 2020).

Research indicates that LMS features can be used to promote SRL processes. The LMS features are significantly effective in activating SRL strategies. Dabbagh and Kitsantas (2005) classifies LMS tool into five groups which include collaborative and communication tools such as discussion forums, chats, and email which can be used to promote goal setting, help-seeking, and time management strategies to learners; content creation and delivery tools such as content view and access, assignment resources, presentations areas and feedback upload features which support learning processes like self-evaluation, task strategies (rehearsal, elaboration, and organization) and goal setting; administrative tools such as tools that help administration of quizzes and tracking of student and course information and calendar that promotes SRL strategies like self-monitoring and help-seeking; assessment tools such as e-portfolios, quizzes and tools that allow for self and peer assessment and finally learning tools such as web links, bookmarking, note-taking, course index, and to-do list features that enhance task strategies.

Table 2.8 presents the LMS features that promote SRL strategies

Table 2.8: Learning Management Systems features that promote SRL Strategies

LMS Tool	LMS Feature	SRL Strategy
Collaborative & communication tools	- discussion forums - chats - email - wikis	- Goal setting - Help-seeking - Time management
Content creation & delivery	- Course & assignment resources - Presentation areas - Feedback uploads - View & access content features,	- Self-evaluation - Goal setting - Task strategies (rehearsal, elaboration, and organization)
Administrative tools	- administer quizzes - tracking and journaling	- Self-monitoring - Help-seeking
Assessment tools	- e-portfolios - quiz tools - peer & self-assessment tools	- Self-evaluation - Self-monitoring
Learning tools	- Search features - weblinks - bookmarking and note-taking - course glossary and index - community & social networking tools	- Task strategies (rehearsal, elaboration, and organization)

(Source: Dabbagh & Kitsantas, 2005)

Dabbagh and Kitsantas (2005) opine that attention should be given to how self-regulatory interventions can be provided especially through LMS features to promote SRL strategies to low self-regulatory learners. There is scanty evidence on how LMS features activate the growth of SRL strategies, especially through artificial intelligence. The existing Educational Data Mining models for SRL focus on MOOCs which are designed and implemented to support learning and teaching for professional and adult learners (Araka et al., 2019). Due to differences in terms of infrastructure and users of MOOCs and LMSs, the findings, tools, and techniques employed on one platform may not be generalized and used on the other platform. There is a lack of empirical evidence on how students regulate on LMSs and how SRL interventions are deployed to learners. Current studies do not address how SRL interventions can be developed especially for LMS coupled with the lack of a suitable model for measuring and scaffolding SRL in e-learning environments (Rowe & Rafferty, 2013).

To realize the potential of LMS features and functionalities in enhancing SRL, empirical research is required to investigate how the use of EDM techniques to provide evaluation and interventions and how they can be deployed within LMS. The nature of online learning requires that students become active participants in their learning process. Since self-regulatory skills are not inherent to students, they need to be developed through learning by students engaging in metacognitive, behavioral, and motivational activities that can foster the growth of SRL (Cerezo et al., 2016); Dabbagh & Kitsantas, 2005).

2.7 Educational Data Mining (EDM)

In this section, we discuss how EDM algorithms are being used to advance the identification and promotion of SRL in online learning environments, the application of clustering and classification algorithms in EDM, the application of agglomerative hierarchical clustering in EDM, and the EDM models used to identify and promote SRL in online learning environments. This gives us an understanding of the SRL strategies to measure within LMS and capture them within the model implemented in this study so that the model takes into account the findings from the previous studies.

Data mining is the process of discovering new knowledge and patterns and building of predictive models from large amount of datasets using computational methods such as machine learning algorithms (Kallio & Tuimala, 2013). Machine learning algorithms are being applied to the dataset from diverse domains such as finance, healthcare, and e-commerce businesses to enrich decision-making by transforming raw data into information (Madni et al., 2017). Educational data mining is also becoming important in analyzing educational data to improve pedagogical aspects of teaching and learning (Coman et al., 2020). EDM emerged over a decade ago and has continued to attract attention from researchers in the computing and education fields. EDM is described as the use of data generated from learning and education environments; applying data mining algorithms to understand learners and improve learning environments. The EDM field also influences the adoption of educational pedagogies established several years ago. Since its emergence, the vision of EDM has been two-fold. First, to use EDM to design and develop learner models that enhance personalization as seen in ITS; Second, to use EDM to enhance learning and teaching theories (Baker, 2016; Baker & Yacef, 2009). The educational data created when learners interact with educational environments can be analyzed to produce inferences to support and improve students' academic performance, used as indicators to provide early interventions, reduce dropout rates and increase retention rates, profile learners, develop learner models, and recommender systems (Arnold & Pistilli, 2012; Romero et al., 2013; Romero & Ventura, 2007).

Figure 2.7 illustrates the various components of EDM that involve the stakeholders who are affected by EDM systems, type of learning environments, sources of the educational data, the EDM tasks, and techniques that can be applied to the educational data (Khanna et al., 2016; Romero & Ventura, 2013; Sin & Muthu, 2015; Sukhija et al., 2016).

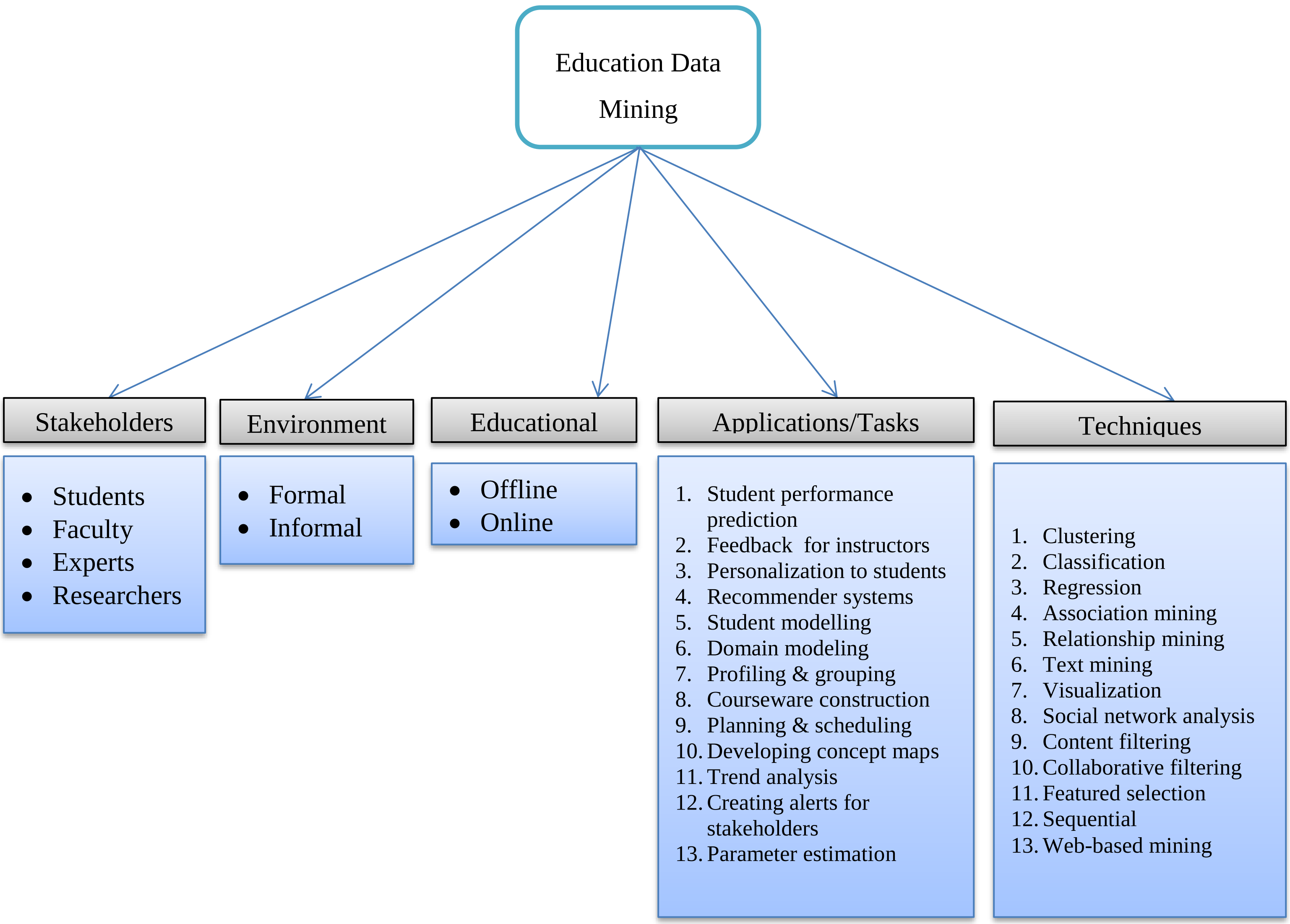


Figure 2.7: Components of Educational Data Mining Systems

According to Winne and Baker (2013), EDM can be used to identify, model, and predict learners' behaviors. Recent studies have applied data mining techniques in educational systems to uncover student behaviors and predict their academic performance (Devasia et al., n.d.; Guo et al., 2015; Oloruntoba & Akinode, 2017; Ratnapala et al., n.d.; Saa, 2016). Various methods of measuring and promoting SRL strategies have started to emerge as a result of the growth of online educational systems. EDM has now made it possible to collect and analyze a variety of datasets from different educational learning environments for analysis to understand learners, tutors, and learning environments. For example, EDM can help us understand how students behave during the learning process concerning self-regulation and to what extent to which the students employ self-regulating skills in e-learning (Kizilcec et al., 2017). EDM has the potential of identifying SRL strategies in real-time and present the results on a dashboard for students and instructors. This information can then be used to determine the nature of support to be offered to students to enhance self-regulatory skills (Panadero et al., 2016). Furthermore, the EDM techniques are likely to identify SRL strategies for online learners more accurately using the actual datasets collected from online learning environments compared to self-report tools.

2.7.1 Educational Data Mining Tasks

The first survey on EDM was carried out by Romero and Ventura (2007) covering the period 1995 to 2005. In the study, two categories of data mining tasks applied in educational data were identified: (a) statistics and visualization and (b) web mining (clustering, classification, outlier detection, association rule mining, sequential pattern mining, and text mining). This survey was followed by another review in 2010 which identified eleven tasks in educational environments that have been resolved through data mining techniques (Romero & Ventura, 2010, 2013). The review by Aldowah et al. (2019) highlights the various EDM techniques that are being applied in solving specific educational problems. After reviewing 402 articles, they suggested four dimensions for classifying EDM based on the educational problems the techniques solve: computer-supported behavioral analytics, computer-supported predictive analytics, computer-supported visualization analytics, and computer-supported learning analytics. Aldowah et al. (2019) focused on computer-supported behavioral analytics in online learning environments. The review by Dutt et al. (2017) concentrated on how clustering data mining techniques have been applied in the context of EDM. The review described the various clustering algorithms that have been used in EDM applications. The systematic review by Papamitsiou and Economides (2014) highlights the key objectives of EDM and LA in educational strategic planning from studies carried out between 2008 and 2013. The study identified six areas of interest where EDM is being applied including student behavior modeling, prediction of performance, increase self-reflection and self-awareness, prediction of dropout and retention, improve feedback and assessment services, and recommendation of resources. The other review by Ganesh and Christy (2015) identified the EDM tasks and analyzed their merits and demerits. They also highlighted the performance results of various EDM techniques applied in solving EDM problems where it was discovered that classification and clustering algorithms were common methods applied to solve EDM tasks. Another review Bakhshinategh et al. (2018) compared existing surveys on EDM tasks and applications and identified 13 categories based on similarity of purpose for each task and scope. The review by

Hernández-Blanco et al. (2019) focused on the application of deep learning machine learning techniques that have been applied in EDM tasks. The authors identified that 4 out tasks have been solved using deep learning approach which includes predicting students' performance, detecting students' undesirable behaviors, providing students' recommendations and automatic evaluations. Table 2.9 presents the summary of the educational data mining applications identified from literature.

Table 2.9: Summary of Educational Data Mining Applications

Reference	Review year	EDM tasks	EDM technique
(Bakhshinategh et al., 2018)	2018	• Adaptive Systems • Evaluation • Scientific Inquiry • Providing Reports • Alert Systems • Planning & Scheduling • Constructing Courseware • Developing Concept maps • Recommender Systems • Performance & Characteristics • Undesirable Behaviors • Profiling & Grouping • Social Network Analysis	• Clustering • Classification • Feature Selection • Outlier Detection • Association Rule Mining
(Ganesh & Christy, 2015)	2015	• Feedback Evaluation • Recommender System • Predicting Performance • Student Modeling • Detecting Activity Behavior • Clustering Students • Social Network Analysis • Schedule Planning • Learning Management System • Dropout & Retention Management • Online Course & E-learning Management	• Clustering • Classification • Visualization • Association rules

(Romero & Ventura, 2013)	2013	• Predicting student performance • Providing feedback for supporting instructors • Personalizing to students • Recommending to students • Creating alerts for stakeholders • Student modeling • Domain modeling • Grouping/profiling students • Constructing courseware • Planning and Scheduling • Parameter estimation	• Prediction • Clustering • Outlier Detection • Relationship Mining • Social Network Analysis • Process Mining • Text Mining • Distillation of Data for Human Judgment • Discovery with Models • Knowledge Tracing • Nonnegative Matrix Factorization
(Romero & Ventura, 2010)	2010	• Analysis and visualization of data • Providing feedback for supporting instructors • Recommending to students • Predicting students' Performance • Student modeling • Detecting undesirable features • Grouping students • Social network analysis • Developing concept maps • Constructing courseware • Planning and Scheduling	No EDM technique

2.7.2 Educational Data Mining Algorithms

Educational data mining involves extracting hidden information about learners and the learning process from educational systems. Through e-learning, large amounts of educational data that can be processed to better understand learners and learning processes is generated (Bogarín et al., 2018a; Romero & Ventura, 2010). EDM provides algorithms and tools that preprocess and analyze educational data to obtain actionable insights about learners and learning environments. Machine learning algorithms can be applied to educational data to explore the data to better understand students and the environments in which they learn. Based on how the machine learning models are trained to perform EDM tasks, the algorithms as applied to educational data are classified into two; supervised learning, and unsupervised learning (Alloghani et al., 2020).

a. Supervised Learning

In supervised learning, labeled data is fed into a machine learning algorithm during the training process. The objective of the training process is to provide the algorithm with the ability to associate certain features with certain outputs (Alloghani et al., 2020; Sathya & Abraham, 2013). Therefore, the result is that the model can learn the features of the data and how they affect the overall outcome. Having achieved this learning, the next step in supervised learning is validation or testing. This is characterized by providing the algorithms with unlabeled data whose true labels, also referred to as target variables, have been set aside. The machine learning algorithm is then expected to predict the labels of this new data based on the knowledge inferred from the training process. Finally, model evaluation in supervised learning is performed when the predictions made from the new test data are compared to the labels that had been set aside during the validation process (Alloghani et al., 2020). Different evaluation metrics exist and these also depend on the type of supervised learning algorithm used.

b. Unsupervised Learning

In unsupervised machine learning, algorithms are tasked with the responsibility of finding patterns within provided data and subsequently use such patterns to categorize data into unique groups based on their inherent similarities (Alloghani et al., 2020; Sathya & Abraham, 2013). Usually, these algorithms are provided with unlabeled data, unlike in supervised learning where the algorithms are provided with labeled data. Using existing mathematical principles and relationships, these algorithms are then able to find similarities and dissimilarities in the provided data and subsequently use them to categorize data into unique classes (Alloghani et al., 2020).

From previous studies, EDM algorithms can be classified into clustering algorithms, temporal data mining, and other techniques such as include natural language processing (NLP) and classification. The techniques used in clustering learners into groups have been achieved using step-wise cluster analysis (Ainscough et al., 2019; Çebi & Güyer, 2020; Valle et al., 2008; Yot-Domínguez & Marcelo, 2017), k-means clustering algorithm (Li et al., 2018), latent class analysis (Barnard et al., 2010), agglomerative hierarchical clustering (Gašević et al., 2017). Literature reveals that among these EDM techniques, clustering algorithms are commonly used.

Table 2.10 presents the summary of the algorithms used to identify SRL in online learning environments (Araka et al., 2022).

Table 2.10: Algorithms used to Measure SRL in Online Learning Environments

	Reference	Data Source	Feature set	EDM Technique	Algorithm used
1.	(Bouchet et al., 2013a)	MetaTutor Trace data & self-report data	• Pageviews • Page visits • Note-taking duration • Session duration • Assessment scores • No. of quizzes completed	Clustering	Expectation-Maximization
2.	(Zheng et al., 2020)	Trace data	• Structural views • Functional shows • Design additions/edits • Note-taking	Clustering	K-means
3.	(Valdiviezo et al., 2013)	LMS trace data	• Course hits • Course views • Assignment views • Forum events • Resources views • Message events • Quiz events	Clustering	K-means
4.	(Maldonado-Mahauad et al., 2018)	MOOC Trace data & self-report data	• Video views • Video reviews • Assessment trials • Course completion status • Assessment reviews • Assessment passes	Clustering	Agglomerative Hierarchical
5.	(Manzanares et al., 2017)	LMS Trace data & self-report data	• Access to course materials • Access to assessments • Access to teacher-feedback • Forum participation • Mean access rates per day	Clustering	Expectation-Maximization
6.	(Sun et al., 2016)	LMS Trace data & self-report data	• Number of assessment attempts • Assessment scores • Time spent on each online lecture • Lecture completion status	Clustering	Agglomerative Hierarchical

7.	(Cicchinelli et al., 2018)	LMS Trace data	• View content indices • View course organization • View exercises • Solve quizzes • View content	Clustering	Agglomerative Hierarchical
8.	(Park et al., 2018)	LMS Trace data	• Video clicks • Quiz submissions • Assignment submissions	Clustering	Probability model-based clustering (Poisson mixture model)
9.	(Kizilcec et al., 2013)	MOOC Trace data	• Forum activity • In-video assessments • Demographic features	Clustering	K-means
10.	(Matcha et al., 2019)	Trace data	• Videos with Multiple-choice questions (MCQs) • Reading materials with MCQs • Exercises	Clustering & Temporal Data Mining	Agglomerative Hierarchical & Expectation-Maximization, Process & Sequence Mining
11.	(Yu et al., 2018)	LMS Trace data	• Video navigations • Assignment views • Quiz views • Discussion sessions	Temporal Data Mining	Neural Networks (LSTM, RNN & GRU)
12.	(Rodriguez et al., 2014)	PLE Trace data	• Blogs • Video Annotations • Bookmarks • Tags • Comments • Excerpts	Temporal Data Mining	Process Mining
13.	(Wong et al., 2019)	MOOC Trace data	• Video views • Quizzes • Assignments • Forum discussions	Temporal Data Mining	Sequential Pattern Mining using Equivalence Classes
14.	(Kinnebrew et al., 2013)	Betty's Brain System trace data	• Reading • Editing • Querying • Explaining • Quizzing	Temporal Data Mining	Differential Sequence Mining
15.	(Cerezo et al., 2020)	LMS trace data	• Forum discussion • Quiz • Resources views • URL views • Course performance	Temporal Data Mining	Inductive Miner

16.	(Di Mitri et al., 2016)	Multimodal data	• Heart rate • Step count • Weather condition • Learning activity	Classification	Regression Analysis
17.	(Syuhada et al., 2020)	Trace data	• Features not mentioned	Classification	K-Nearest Neighbor
18.	(Bosch et al., 2018)	LMS Trace data	• Total logins • No. of events per login • Total interaction events • Access to materials • Grade views • Quiz attempts • Correct quiz answers • Exam attempts • Correct exam attempts • Forum post views • Forum posts created	Classification	Logistic regression
19.	(Rodriguez et al., 2019)	LMS Trace data & self-report data	• Video clicks • Slide clicks	Statistical modeling	Binomial Regression
20.	(Montgomery et al., 2019)	LMS Trace data	• Access location • Access time (of the day) • Online login frequency • Online login regularity • Quiz review pattern • Course materials views	Statistical modeling	Association & Correlational analysis
21.	(Trevors et al., 2016)	Multimodal data & Self-report data	• Eye-tracking patterns • Study tools • Metacognitive ratings	Statistical modeling	Correlation analysis
22.	(Jansen et al., 2020)	MOOC Trace data & self-report data	• Video interaction events • Quiz interaction events • Marking reading as completed • Submission of assignment • Page navigations • Visits & posts on forums	Statistical modeling	Statistical modeling
23.	(Jo et al., 2016)	LMS trace data & self-report data	• Login frequency • Login regularity • Total login time	Statistical modeling	Statistical modeling
24.	(Crossley et al., 2016)	MOOC Trace data	• Video interaction • Forum interaction • Pageviews/Assignments	Natural Language Processing (NLP) tools	WAT, TAALES, TAACO, ReaderBench & SEANCE

Since online learners' behaviors differ from student to student, it is necessary to use clustering algorithms to identify the groups in which the learners belong to provide targeted SRL interventions compared to offering one-size-fits-all interventions to students.

2.7.3 Clustering Algorithms

Clustering algorithms belong to unsupervised machine learning algorithms that identify data points in unlabeled datasets and isolate them into groups (clusters) based on the similarity and dissimilarity of the data points such that a cluster contains data points that are similar to each other but dissimilar to data points in another cluster. The algorithms are classified into four categories namely connectivity, partition, centroid, and density-based algorithms (Halkidi et al., 2001). Clustering students based on their learning behaviors provide insights into understanding how students learn and hence an opportunity to design interventions targeted to especially students with undesirable characteristics (Araka et al., 2022). This section describes the characteristics and operation of the frequently used clustering algorithms namely k-means, expectation-maximization, and agglomerative clustering algorithms (Araka et al., 2022). As presented in Table 2.10 machine learning algorithms can be applied to different EDM applications such as classification, clustering, association rules, and recommender systems that utilize educational data to provide insights that help enhance the understanding of learners and learning environments. Of these tasks, clustering is the most frequently and commonly used in educational data mining. The k-means, expectation-maximization and agglomerative hierarchical clustering have frequently been applied to educational data (Araka et al., 2022; Elsayed et al., 2019; Job, 2018; Valarmathy & Krishnaveni, 2019).

2.7.3.1 K-means

The k-means algorithm is a centroid-based algorithm that forms clusters based on the density (space) between data points. The k-means algorithm requires prior specification of the number of clusters, k (Naeem & Wumaier, 2018). Initially, k initial centroids are chosen, with k being an arbitrary choice of the number of clusters that the data will be split into. The k-means algorithm then determines the centroid for each cluster assigning to itself the point which is closest to the centroid (Alloghani et al., 2020; Mouton et al., 2020). Every individual data point in the provided dataset is then allocated to the closest centroid. A collection of related data points around a given centroid is then considered to be a cluster. A data point is assigned to a cluster that has a mean nearest to the data point. Once the data point is added, the mean for that cluster is updated together with the distance between the centroid of the cluster. The algorithm, therefore, keeps track of the cluster means of the same feature space as x. The algorithm, therefore, works well with data that is circular and spherical. The random choice of k may give different clusters in each run. This implies that the clusters may not be reproduced and hence lack consistency. The k-means has a linear complexity O(n) and hence can handle large datasets.

The pseudo-code is as follows:

1. Randomly choose k centers μ_1, μ_2, …μ_k
2. Assign x_1, x_2, …, x_n to their nearest centroids
3. Update μ_i to the mean of its data points.
4. Repeat 1, 2, and 3 until there is no change in the clusters.

2.7.3.2 Expectation-Maximization

Expectation-Maximization (EM) is a distribution-based algorithm that is grounded on the knowledge that data points in a cluster have a probability of being in the same distribution; Gaussian or normal. The algorithm determines the probabilities of data points in each cluster. The algorithm operates in two steps; the expectation (E) and the maximization (M). In each step, the means and standard deviations for each cluster are computed to maximize the probability of the distributions of the data points. The data points are therefore assigned to clusters based on probability. This renders EM more flexibility regarding cluster covariance compared K-means algorithm by taking any dataset that may many not be circular. The EM algorithm has a linear complexity O(n) in terms of time.

The pseudo-code is as follows:

1. Randomly pick initial parameters for mean and standard deviation.
2. Repeat; refine the parameters in E and M steps.
3. In the E step, compute the cluster for each data point based on the initial mean and standard deviation.
4. In the M step, re-compute the parameters based on the clusters where data points have been placed guided by probability
5. Each data point is assigned to a cluster in which it has the highest probability.

2.7.3.3 Agglomerative Hierarchical Clustering

The agglomerative hierarchical clustering algorithm is a connectivity-based that groups data points into their clusters and then merges the clusters progressively until one cluster with all the data points is obtained. (Mouton et al., 2020). This process obtains a hierarchical tree referred to as the dendrogram which signifies the distance between the aggregated clusters. The number of clusters in the dataset is obtained by measuring the maximum vertical distance between the nodes on the dendrogram. The Algorithm has a time complexity of $O(n^3)$ and memory complexity of $O(n^2)$. The algorithm however is not efficient for large datasets.

The pseudo-code is as follows;

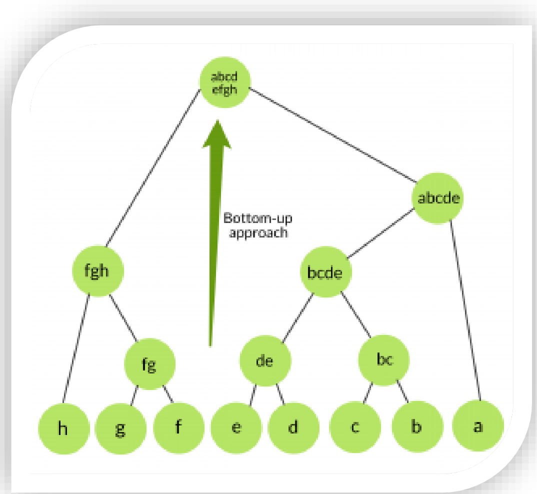
1. Initially, each data point $x_1, x_2, \ldots x_n$ is in its cluster $c_1, c_2, \ldots, c_n$
2. Repeat progressively until only one cluster is formed.
3. Aggregate the nearest clusters, say C_i and C_j.
4. Output: *dendrogram* that is used to obtain the optimal number of clusters

Agglomerative hierarchical clustering aggregates the distance between data points into the distance between clusters using four strategies namely single linkage, complete linkage, average linkage, and centroid methods (Schubert, 2020). The limitation of AHC is the requirement to compute and store a full distance matrix that leads to the time and space complexity of O(n2) on the lower bound and O(n3) on the upper bound. Besides the empirical evaluation results that indicate agglomerative hierarchical clustering as the optimal algorithm with the optimal number of clusters (Araka et al., 2022), the characteristics and operation of the algorithm indicate that it is best suited for educational data in higher institutions of learning. The nature of student behavioral data is that it evolves during a semester. The resulting clusters formed as time changes represent a hierarchical tree-like structure. When data has a hierarchical structure and the objective is to discover the underlying cluster trends over time as in the case of the present study, then the agglomerative hierarchical clustering algorithm is the best choice (McBroom et al., 2020).

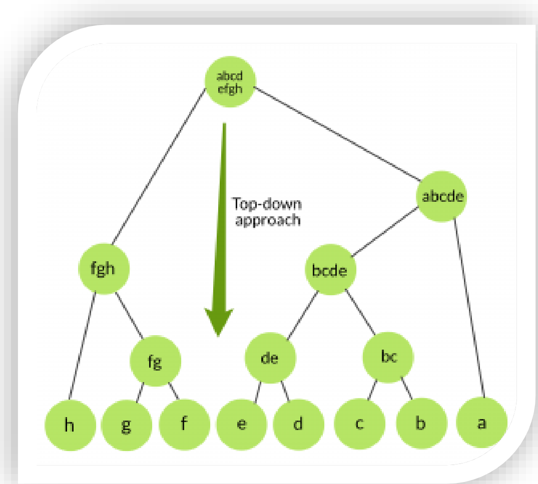
2.7.4 Application of Agglomerative Hierarchical Clustering for EDM

This section discusses the application agglomerative hierarchical clustering algorithm to educational data. Hierarchical clustering can be categorized into two: agglomerative hierarchical clustering (bottom-up) where all data points are placed in individual clusters and then iteratively merged with closest clusters until one cluster is formed and divisive hierarchical clustering (top-down) which initially starts with one cluster that is recursively split until we have each data point on its cluster (Charrad et al., 2014; Halkidi et al., 2001).

In this section agglomerative hierarchical clustering technique which measures the number of clusters based on the distance between clusters is highlighted since it is the commonly used approach for hierarchical clustering (Contreras & Murtagh, 2015). Additionally, agglomerative hierarchical clustering provides internal and constituent structures on clusters that are formed unlike other clustering techniques (Aljumily, 2017). The agglomerative hierarchical clustering algorithm has also been recognized as a suitable algorithm for identifying students' behavioral changes over time during learning in online environments (McBroom et al., 2020). Figure 2.8 illustrates the divisive and agglomerative categories of the hierarchal clustering algorithm.



(a) Agglomerative hierarchical clustering



(b) Divisive hierarchical clustering

Figure 2.8: Divisive and Agglomerative Hierarchical Clustering Algorithms

Note. Approaches to hierarchical clustering: (a) agglomerative hierarchical clustering and (b) divisive hierarchical clustering.

According to McBroom et al. (2020), agglomerative hierarchical clustering can be able to detect behavioral trends over time unlike the standard clustering algorithms such as the k-means, k-nearest neighbor, and expectation-maximization, especially when applying the algorithms to time series data which is temporal. For instance, k-means that seek to minimize the distance between data points only group students with similar learning behaviors. As a result, k-means may not be able to work well with time series data where many features may not be related because of the changes in trends in students learning patterns. To explore the implementation of agglomerative hierarchical clustering on educational data, a case study was carried out on the Open University Learning Analytics Dataset (OULAD). The dataset obtained from the Open University in the UK represents learners' interactions in an online learning environment inform of the number of interactions (visits) with various learning resources and learning activities students engaged in (Araka et al., 2022; Jha et al., 2019; Kuzilek et al., 2017). Table 2.11 presents a sample of the OULAD dataset that was used for the case study.

Table 2.11: Sample of the Open University Learning Analytics Dataset used in the Case Study

Semester	Student ID	Number of visits	Final Result
2013J	123044	2041	Pass
2013J	127582	1321	Pass
2013J	129955	1011	Withdrawn
2013J	132976	520	Pass
2013J	134143	2732	Distinction
2013J	1352868	641	Withdrawn
2013J	135335	57	Withdrawn
2013J	135400	287	Withdrawn
2013J	137873	5528	Pass
2013J	1401935	240	Fail
2013J	1402638	910	Pass
2013J	141355	1823	Pass

To establish the suitability of the Open University learning analytics dataset in formulating separated clusters, a pair plot was visualized using ***sum_click*** and ***id_student*** variables.

Figure 2.9 illustrates the visualized pair plot.

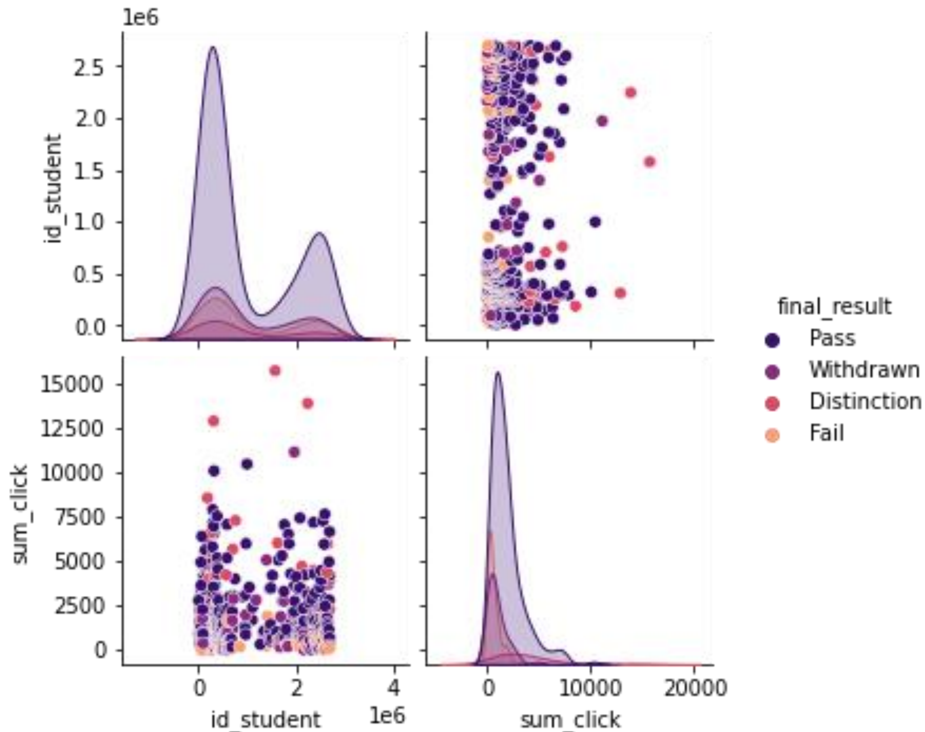


Figure 2.9: Pair Plot for the OULAD dataset

The modeling phase involved exploring various distance measures and linkage strategies to identify the appropriate distance measure and linkage strategy that can be applied to educational data. The internal validity metrics that were used to evaluate the quality of the clusters formed included the Silhouette score, the Calinski-Harabasx index & the Davies-Bouldin index (Charrad et al., 2014; Halkidi et al., 2001; Kumar & Raju, 2018; Liu et al., 2010). The Silhouette score is used to measure the separation distance between clusters and ranges between -1 and 1. The closer the score is to one, the further a cluster's samples are far from the neighboring clusters' samples (Charrad et al., 2014; Kumar & Raju, 2018). The Silhouette score is computed using the mean of intra-cluster distance, *i*, and the mean inter-cluster distance, *n*, for every sample by the formula;

(n - i) / max (i, n).. (1)

The Calinski-Harabasz index is based on the knowledge that clusters that are themselves very compact and well-spaced from each other are good. The Calinski-Harabasz index is computed by dividing the variance of the sums of squares of the distances of individual objects to their cluster center by the sum of squares of the distance between the cluster centers (Kumar & Raju, 2018; Liu et al., 2010). The higher the Calinski-Harabasz index, the optimal the clustering model. The Calinski-Harabasz index is given by the formula;

$$s = \frac{\mathrm{tr}(B_k)}{\mathrm{tr}(W_k)} \times \frac{n_E - k}{k - 1} \qquad (2)$$

Where tr(Bk) is the trace of the between-group dispersion matrix and tr(Wk) is the trace of the within-cluster dispersion matrix which is defined by the formulae;

$$W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - c_q)(x - c_q)^T \qquad (3)$$

$$B_k = \sum_{q=1}^{k} n_q (c_q - c_E)(c_q - c_E)^T \qquad (4)$$

Where c_q is the set of points in cluster q, c_E is the center of E, and n_q is the number of points in cluster q.

The Davies-Bouldin index, like the silhouette score and the Calinski-Harabasz index, captures both cluster separation and compactness. The index is defined as the average similarity measure of each cluster with its most similar cluster, where similarity is the ratio of within-cluster distances to between-cluster distances (Davies & Bouldin, 1979; Kumar & Raju, 2018). Hence, unlike the silhouette score and the Calinski-Harabasz index, as the Davies-Bouldin index decreases, the clustering improves (Liu et al., 2010). The Davies-Bouldin formula is given by (Charrad et al., 2014);

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} R_{ij} \qquad (5)$$

Where R_{ij} *is defined by*;

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \qquad (6)$$

Where;

(i) s_i is the cluster distance given by the average distance between each point of cluster *i* and the centroid of that cluster and

(ii) d_{ij} is the distance between cluster centroids i and j.

Distance measures are used to identify the closeness and similarity of data points and cluster them together. To explore the distance between data points (in clusters), four distance measures were considered in the case study (Charrad et al., 2014).

1) Euclidean distance measure technique computes the distance between data points by quantifying two data points using Pythagoras Theorem and is defined by the following formula;

   D (Euclidean) = $\sqrt{[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]}$... (7)

2) Mahalanobis distance which is used to compute the distance between two data points in terms of t-score is defined by the formula;

   D (Mahalanobis) = $\sqrt{[(\mathbf{x} - \mathbf{y})^T * C^{-1} * (\mathbf{x} - \mathbf{y})]}$.. (8)

3) Minkowski distance is defined by the formula:

   D (Minkowski) = $(\sum|x-y|^p)^{1/p}$.. (9)

Where;

(i) x and y are variables (data points).
(ii) C is a covariant matrix of the data containing x and y.
(iii) p is a constant where if p=1, the Minkowski distance is similar to the Manhattan distance, and if p=2, the Minkowski distance is similar to the Euclidean distance.

Besides the distance measures, linkage strategies are used in agglomerative hierarchical clustering to merge clusters to form bigger clusters. The following linkage strategies were considered in the case study (Charrad et al., 2014);

1. Single linkage strategy that merges clusters based on the shortest distance between the nearest clusters (neighbor). This strategy is not preferred as it may result in prematurely aggregated clusters.

2. Complete linkage strategy that merges the clusters based on the longest distance between possible pairs of nearest clusters resulting in more compact clusters.
3. The average linkage strategy computes the average distance between the neighbor cluster and lastly.
4. The centroid linkage method considers the distance between cluster centers.
5. The ward's linkage strategy computes the distance between clusters and merges the clusters based on the sum of the squares index.

The output from the implementation of the distance measures and linkage strategies as applied to the Open University learning analytics dataset is presented in Figures 2.10 – 2.19.

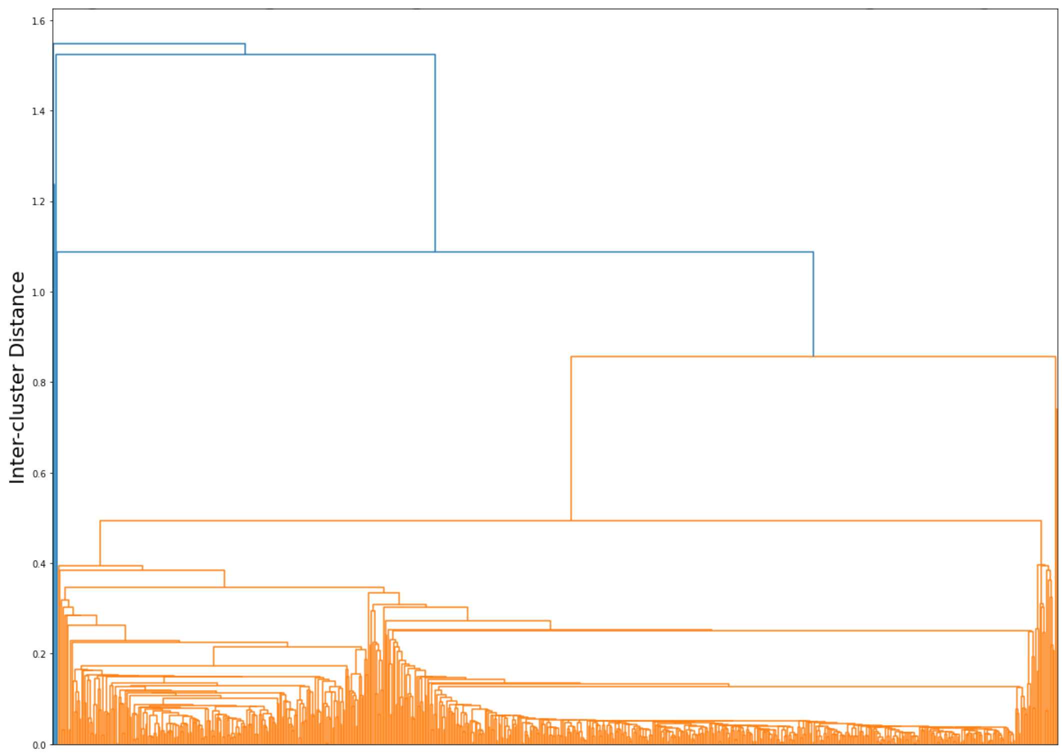


Figure 2.10: Dendrogram showing clustering with Euclidean distance and Single Linkage

Note. The Silhouette score is 0.73, the Calinski-Harabasz score is 40.76, and the Davies-Bouldin score is 0.29.

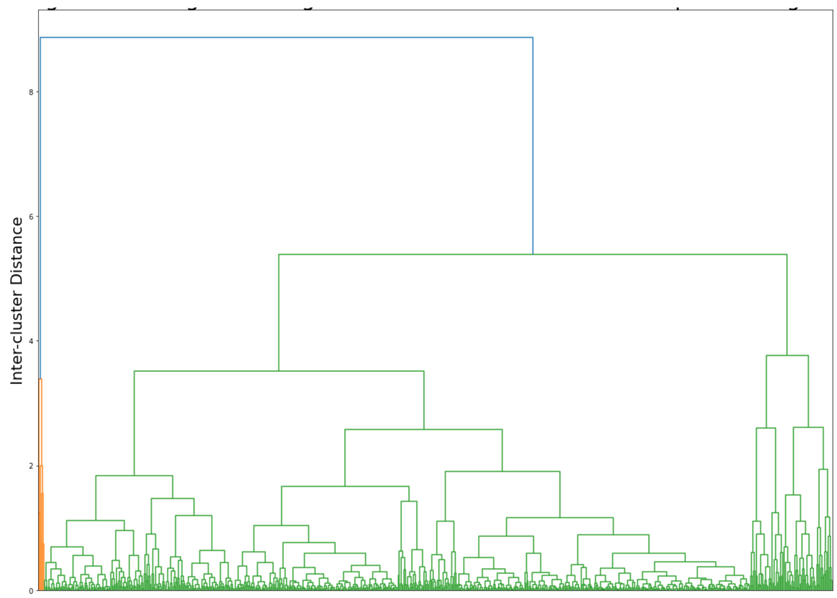


Figure 2.11: Dendrogram showing Clustering with Euclidean distance and complete linkage

Note. The Silhouette score is 0.48, the Calinski-Harabasz score is 219.95, and the Davies-Bouldin score is 0.87.

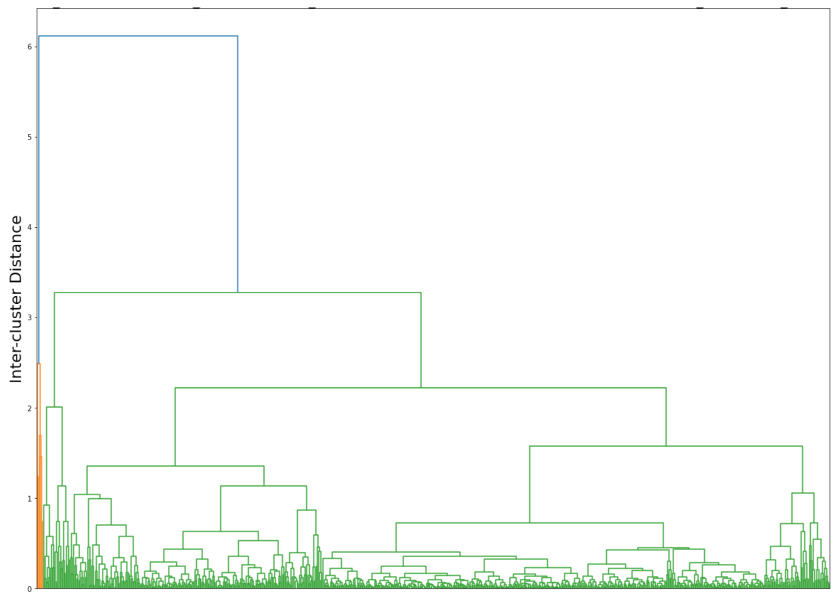


Figure 2.12: Dendrogram showing Clustering with Euclidean distance and Average linkage

Note. The Silhouette score is 0.53, the Calinski-Harabasz score is 147.82, and the Davies-Bouldin score is 0.75.

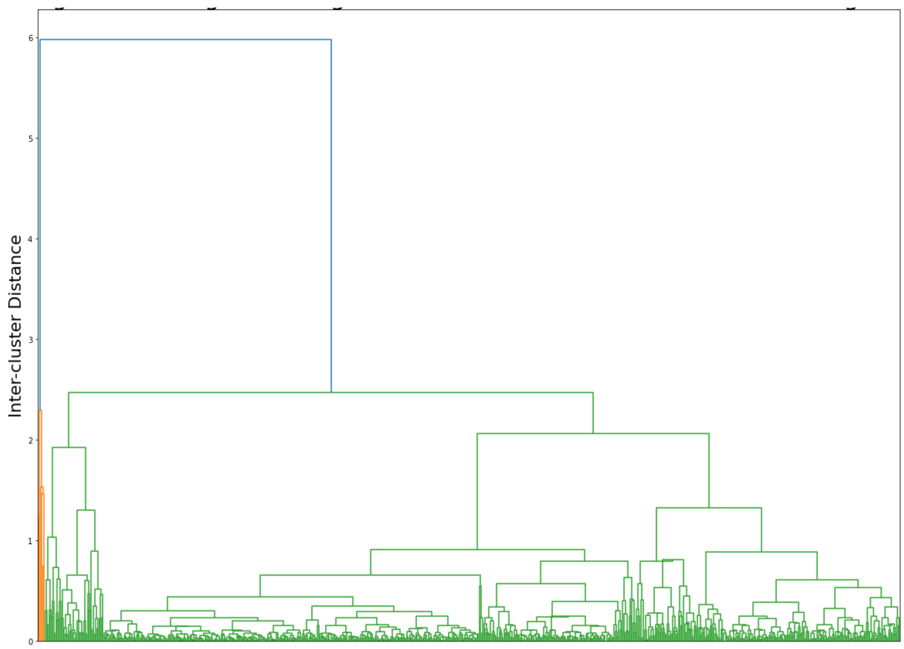


Figure 2.13: Dendrogram showing Clustering with Euclidean distance and Centroid linkage

Note. The Silhouette score is 0.49, the Calinski-Harabasz score is 186.43, and the Davies-Bouldin score is 0.78.

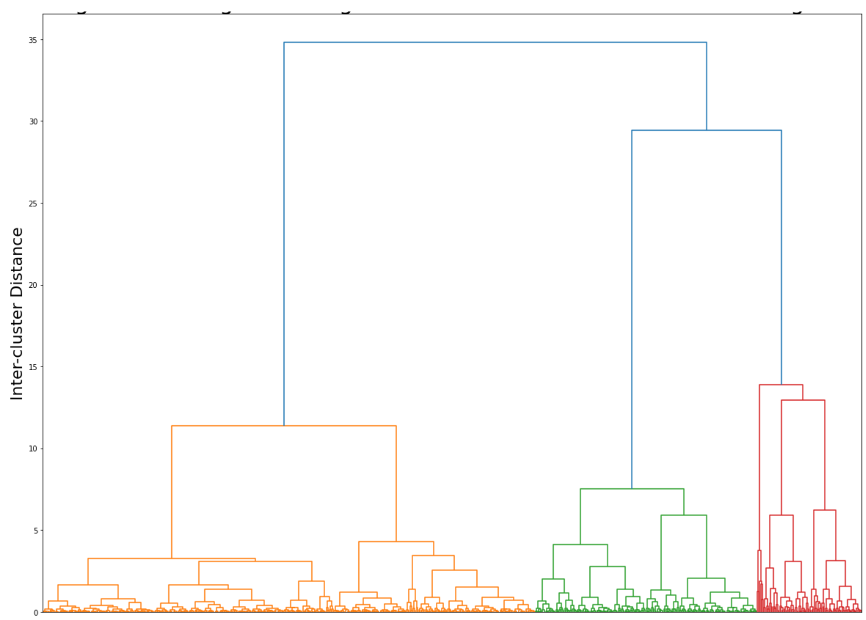


Figure 2.14: Dendrogram showing Clustering with Euclidean distance and Ward linkage

Note. The Silhouette score is 0.61, the Calinski-Harabasz score is 880.38, and the Davies-Bouldin score is 0.75.

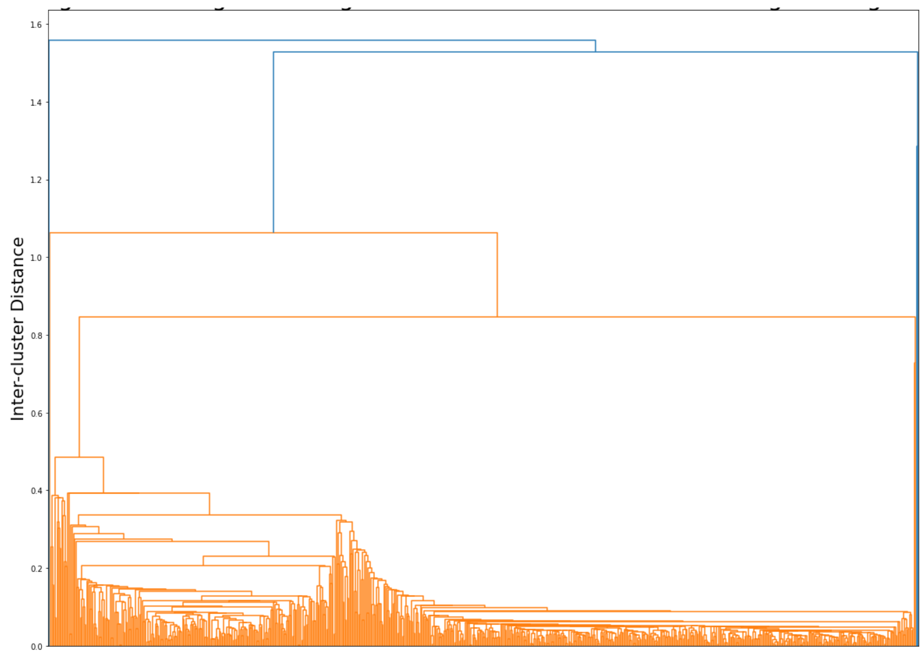


Figure 2.15: Dendrogram showing Clustering with Mahalanobis distance and Single linkage

Note. The Silhouette score is 0.73, the Calinski-Harabasz score is 40.76, and the Davies-Bouldin score is 0.29

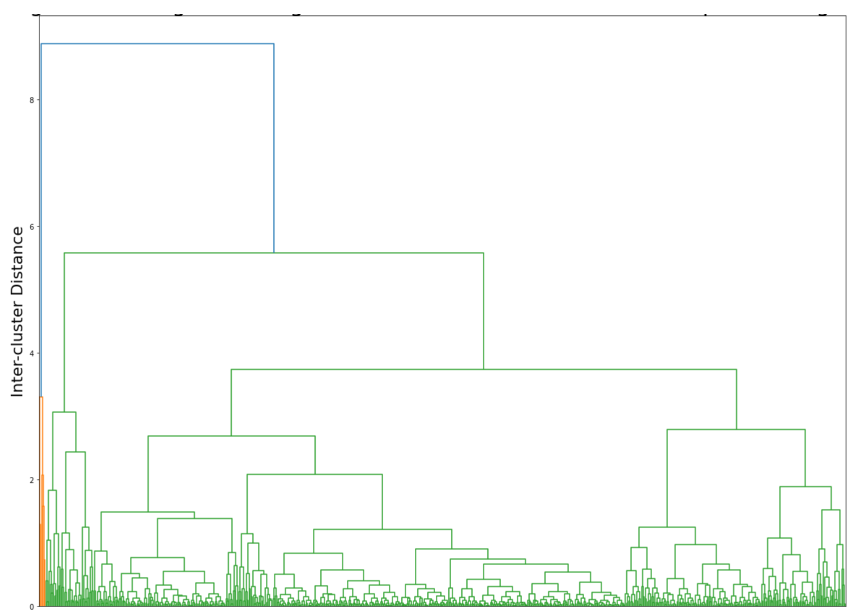


Figure 2.16: Dendrogram showing Clustering with Mahalanobis distance and Complete linkage

Note. The Silhouette score is 0.51, the Calinski-Harabasz score is 181.47, and the Davies-Bouldin score is 0.76.

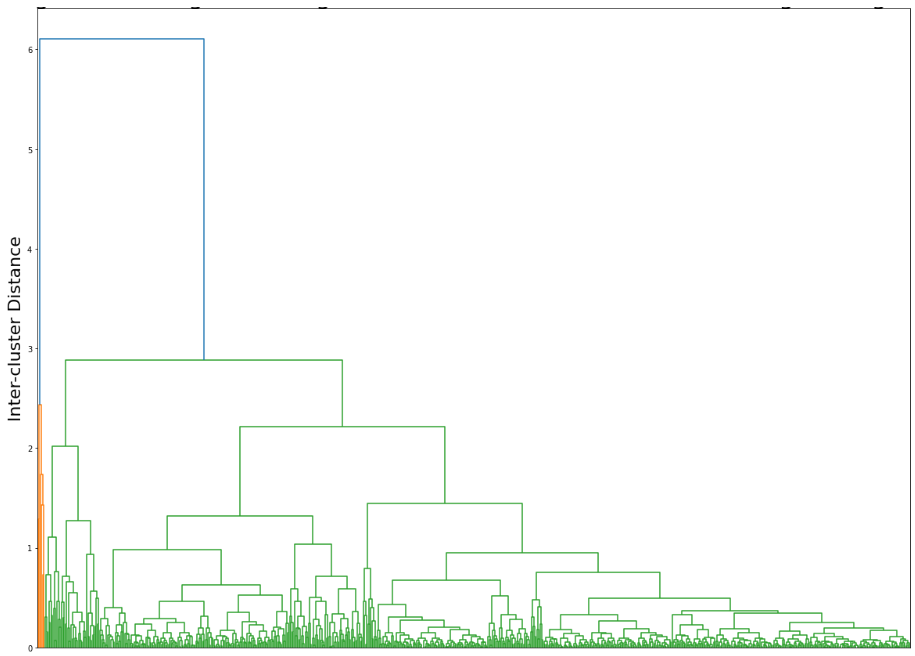


Figure 2.17: Dendrogram showing Clustering with Mahalanobis distance and Average linkage

Note. The Silhouette score is 0.50, the Calinski-Harabasz score is 183.07, and the Davies-Bouldin score is 0.77.

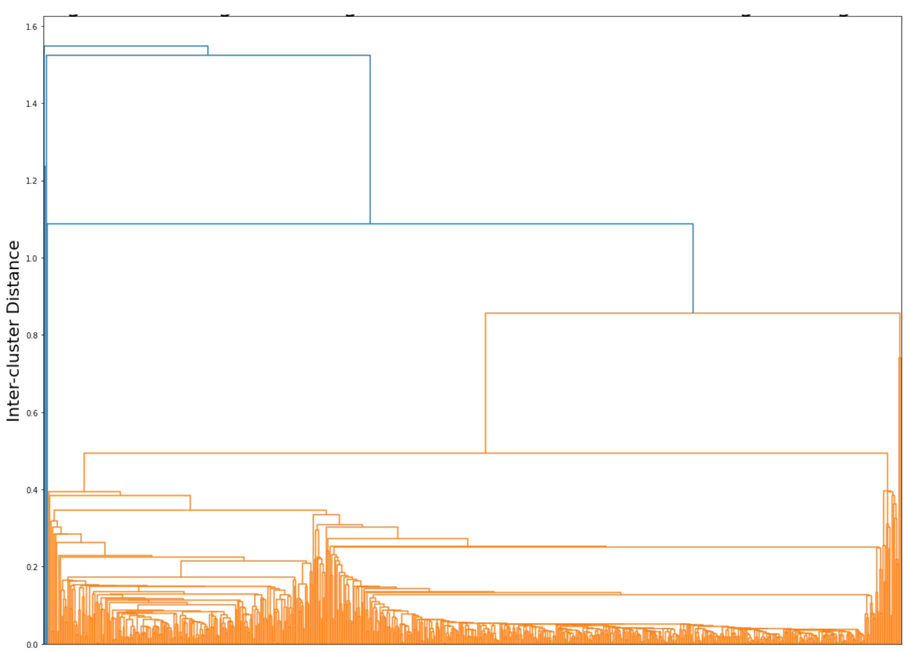


Figure 2.18: Dendrogram showing Clustering with Minkowski distance and Single linkage

Note. The Silhouette score is 0.73, the Calinski-Harabasz score is 40.76, and the Davies-Bouldin score is 0.29

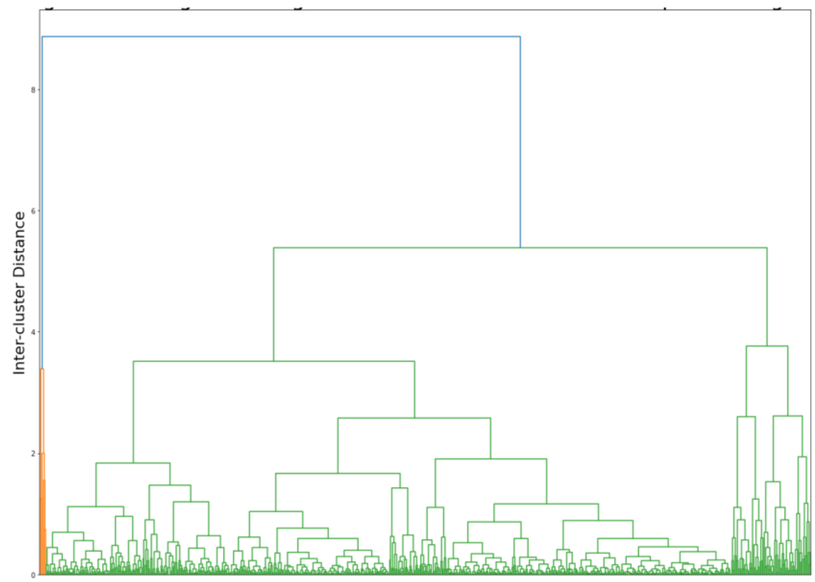


Figure 2.19: Dendrogram showing Clustering with Minkowski distance and Complete linkage

Note. The Silhouette score is 0.48, the Calinski-Harabasz score is 219.95, and the Davies-Bouldin score is 0.87.

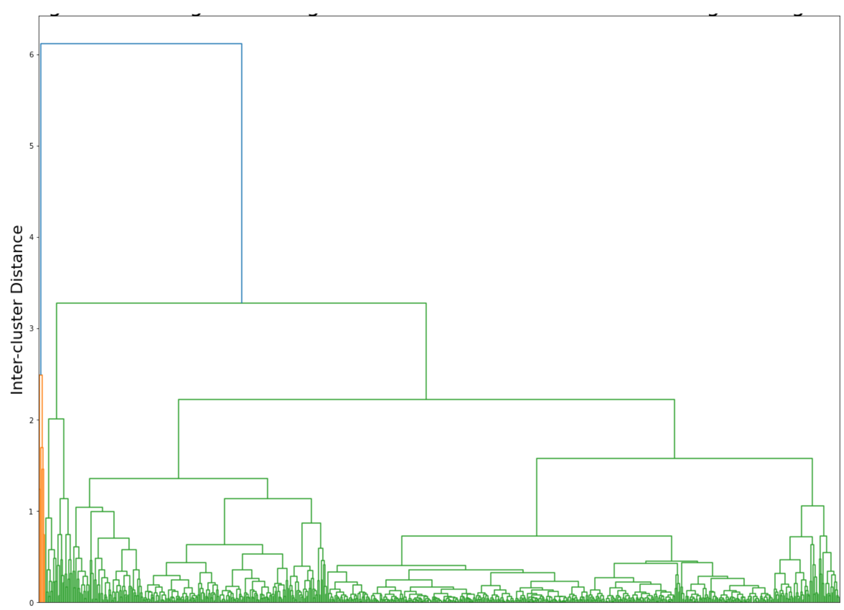


Figure 2.20: Dendrogram showing Clustering with Minkowski distance and Average linkage

Note. The Silhouette score is 0.53, the Calinski-Harabasz score is 147.82, and the Davies-Bouldin score is 0.75.

During the implementation of the above distance measures and linkage strategies, internal validity measures were computed for each distance and linkage strategy combination and a summary is presented in Table 2.12.

Table 2.12: Internal Validity Scores for various Distance measures and linkage methods

Distance measure	Linkage method	Internal Validity Score		
		Silhouette	Calinski-Harabasx	Davies-Bouldin
Euclidean	Single linkage	0.73	40.76	0.29
	Complete linkage	0.48	219.95	0.87
	Average linkage	0.53	147.81	0.74
	Centroid linkage	0.49	186.43	0.78
	Ward linkage	0.61	880.38	0.75
Mahalanobis	Single linkage	0.73	40.76	0.29
	Complete linkage	0.51	181.47	0.76
	Average linkage	0.50	183.07	0.77
Minkowski	Single linkage	0.73	40.76	0.29
	Complete linkage	0.48	219.95	0.87
	Average linkage	0.53	147.82	0.75

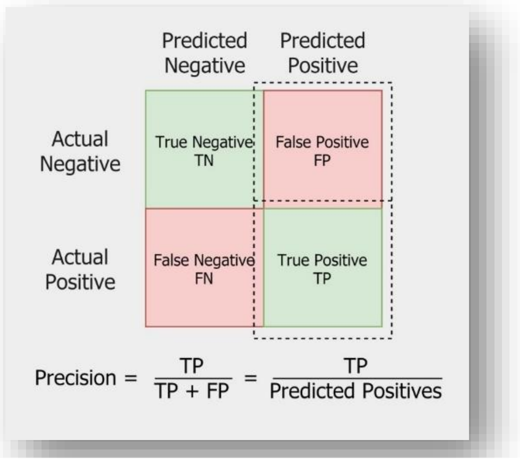
From the results of the application of the agglomerative hierarchical clustering algorithm on the educational data the following was observed;

a) When the linkage strategies are compared with one another, the best performing method is the single linkage method which returns a silhouette score of 0.73. The worst performing method is the complete method which returns a silhouette score of 0.48.
b) When linkage strategies are combined with different distance measures, the silhouette score is the same for all distances while using the single method with a score of 0.73. When using the complete linkage method, the Mahalanobis distance performs best, with a score of 0.5. When using the average linkage method, the Minkowski and Euclidean distances perform best, with a score of 0.53.
c) The overall best-performing clustering model is the hierarchical agglomerative clustering algorithm with any distance while using the single linkage method.

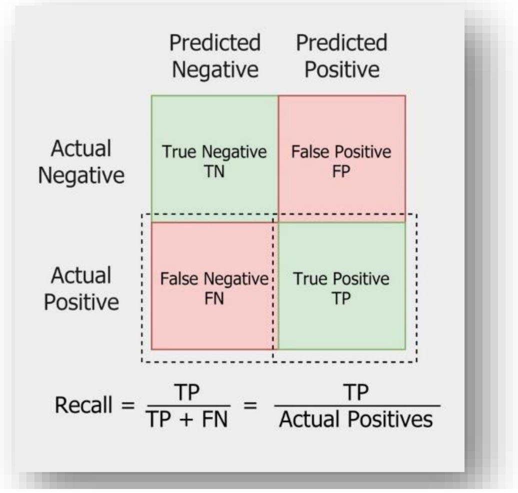
2.7.5 Classification Algorithms

This section describes the characteristics and operation of the classification algorithms that were implemented in the development of the EDM model. The performance of the algorithms was also evaluated during the EDM plugin development and the optimal classifier was selected and implemented in the development of the EDM model. The algorithms include logistic regression, decision tree, random forest, and XGBoost classifier. Classification refers to supervised learning tasks where a machine learning algorithm is expected to predict labels belonging to a given category. Given the nature of the output of the algorithm, classification algorithms are evaluated using a unique set of evaluation metrics as compared to regression algorithms. Commonly used evaluation metrics for classification tasks include accuracy, F_1 score, precision, recall, and log loss. Again, the choice of metric in a classification task is highly dependent on the characteristics of the data used for the tasks. It is common to find observations in data belonging dominantly to one category, while the other categories remain significantly underrepresented. These occurrences are common in tasks such as cancer detection, fraud detection, and intrusion detection (Alloghani et al., 2020). The imbalance can be explained by the fact that the occurrence of cancer, fraud, or an intrusion attack in a computer network happens less frequently and therefore most of the data will belong to normal tissues, no fraud, and normal intrusion in the three cases respectively (Louppe, 2014). Given this rationale, using a metric like accuracy would provide an indicator of the probability of predicting the dominant class due to the high imbalance. To overcome this challenge precision and recall metrics are used. Figure 2.21 shows the computation of precision and recall in a classification problem using a confusion matrix. After the computation of precision and recall, the F_1 score is then given by the formula;

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{tp}}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})}.$$



(a) Computation of precision in a classification problem



(b) Computation of recall in a classification problem

Figure 2.21: Computation of precision and recall in a classification problem

Note. Precision focuses on the number of positive predictions made by the classifier that has been trained. Recall returns the proportion of actual positives that were correctly predicted by the algorithm. The use of precision and recall as evaluation metrics is subjective based on the trade-off associated with a two. High precision results in a lower recall while the reverse is true.

2.7.5.1 Logistic Regression

Logistic regression is a classification algorithm used in machine learning tasks. Although the model is considered a linear model, its principle of operation is based on the sigmoid function, thus implying that it is probabilistic. Unlike the linear regression model which returns continuous values, the logistic regression function, also known as the sigmoid function, returns outputs that belong to one of two or more categories (Hyeoun-Ae, 2013; Peng et al., 2002). Logistic regression algorithms can be categorized as either binomial or multinomial regression. In binomial regression, the target category being investigated belongs to one of two categories. In multinomial regression, the target variable belongs to more than one category. Common binomial tasks include email classification as spam and non-spam while typical multinomial tasks include the MNIST hand-written dataset. The sigmoid function is used to compute the probability of an observation belonging to one of many classes (Hyeoun-Ae, 2013; Peng et al., 2002). The function is summarized in the equation:

$$F(x) = 1/(1+e^{-(x)})$$

Given their probabilistic nature, it is important to provide decision boundaries to logistic regression algorithms such that the boundaries serve as thresholds for different classes to be predicted. In binary classification problems, a typical threshold value of 0.5 is used implying that all predictions with probabilities beyond 0.5 will be assigned to the first category. Figure 2.22 shows how the sigmoid function is used to detect attacks on IoT devices that are used to collect training data for machine learning (Baracaldo et al., 2018). The figure illustrates a binary classification problem with highly imbalanced data. Logistic regression models tend to perform better in cases involving imbalanced classes due to the probabilistic nature of the sigmoid function.

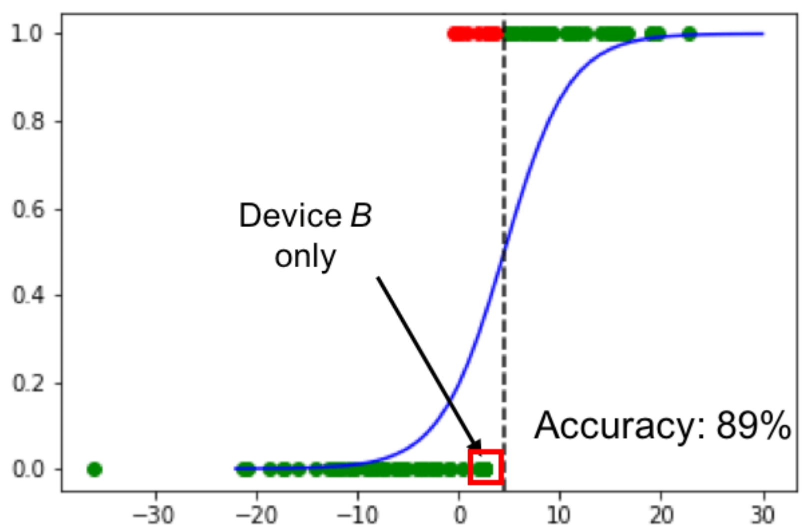


Figure 2.22: Performance of Logistic Regression Model

2.7.5.2 Decision Tree

Decision tree algorithms are a group of tree-based algorithms that use hierarchical tree structures to arrive at decisions given a set of data. These algorithms can be used for both classification and regression tasks and as a result, they are commonly abbreviated as Classification and Regression Trees (CART) (Alloghani et al., 2020). Decision Tree algorithms have found applications in both data mining where they are used to gain inferences from large sets of data and in machine learning tasks where they are used for classification and regression (Rokach & Maimon, 2008). The structure of a decision tree can be thought of as an inverted tree with the root node at the top and the decision or leaf at the bottom. The parent root refers to a particular characteristic of the dataset that can be used to split the data into two broad categories. Other smaller roots are then derived from this parent root until a final decision is achieved at the terminal node or leaf. The size of a tree depends on characteristics such as the features present in the data, conditions to be used in splitting the features, and finally, the limit that would indicate the terminal node. In typical settings, a feature is considered as the parent node if it has the least cost. Typically, the Gini score is used to evaluate the appropriateness of a particular split within the tree, and subsequently, it can be used to obtain the depth of a particular tree. In the diagram below, a decision tree algorithm used in the classification of cardiac arrhythmia is shown. One characteristic of the root node is that it can be used to split the data into two broad categories based on a given threshold. Figure 2.23 illustrates the operation of the decision-tree algorithm that was used to classify Cardiac Arrhythmia using the conditional C_i threshold (Elsayyad et al., 2015). Using the C_i metric, all observations falling below the value are considered to be arrhythmia, while those with values greater than the threshold are subjected to further test conditions.

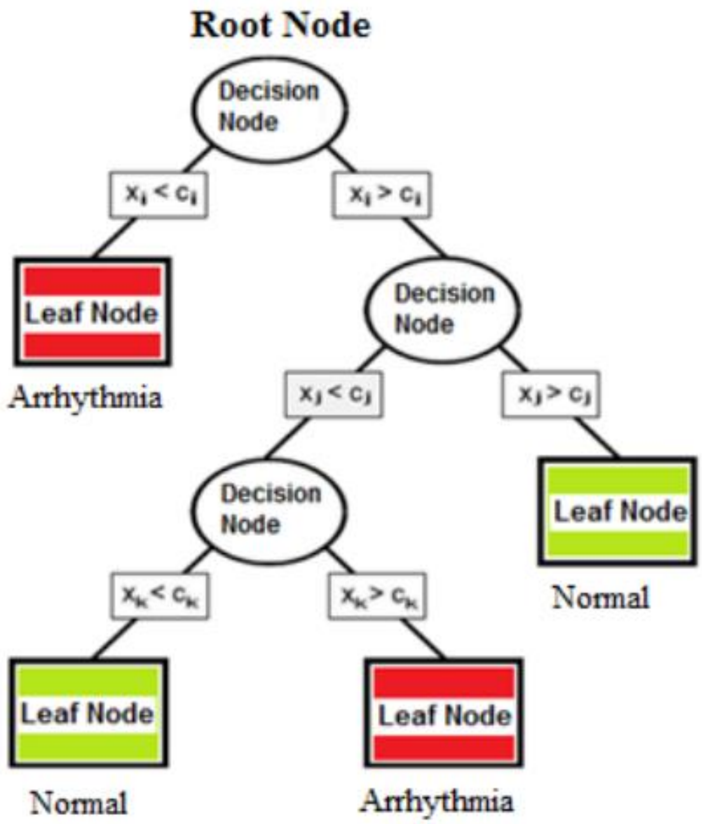


Figure 2.23: Decision-Tree Algorithm used in the Classification of Cardiac Arrhythmia

2.7.5.3 Random Forest

The random forest algorithm leverages on the principle of operation of several decision trees put together horizontally. Based on this principle, the algorithm belongs to a wider class of algorithms known as ensemble models. These are characterized by the fact that they make decisions based on predictions from other smaller classifiers such as decision tree algorithms in a process known as bootstrap aggregating or bagging (Ho, 1995; Louppe, 2014). To come up with a final prediction, random forest algorithms use various techniques. These may include taking the average of the individual predictors from the individual trees and treating the average as the final prediction or in some instances, taking the prediction that occurs the maximum number of times from the ensemble of decision trees as the final prediction. Under their principle of operation, random forest algorithms tend to perform better than standalone decision tree models. The ensemble approach also implies that the random forest algorithm is typically less prone to overfitting as compared to the decision trees. Other reasons why the random forest is a commonly preferred algorithm in classification tasks include its ability to effectively handle missing data properly, its effectiveness when handling large datasets, and the algorithm's ability to work effectively with little hyper-parameter tuning. On the downside, the algorithm works poorly when given extrapolation tasks, a possible indicator of why it is used less frequently in regression tasks (Louppe, 2014). Figure 2.24 illustrates how a random forest algorithm was applied in the classification of spacecraft electrical signals (Li et al., 2017). It can be observed that the maximum vote approach is seen to be utilized where predictions from the individual bootstrap sets are considered as individual votes and thereafter the predictions with the most number of occurrences are considered the maximum vote and hence the final random forest prediction.

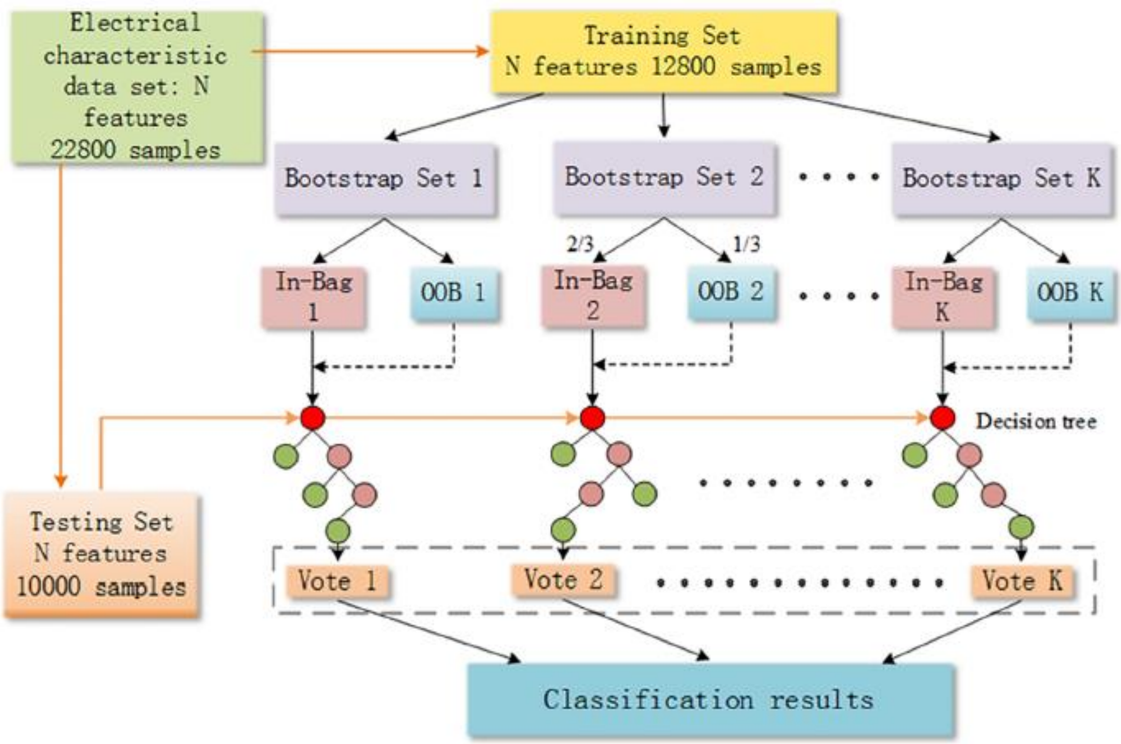


Figure 2.24: Random Forest Algorithm used to Classify the Spacecraft Electrical Signals

2.7.5.4 Extreme Gradient Boosting

The extreme gradient boosting (XGBoost) algorithm is a tree-based algorithm that also belongs to the ensemble category of models. Unlike the random forest algorithm that relies on bootstrap aggregating to make predictions, the XGBoost algorithm leverages on gradient boosting to improve model performance (Chen et al., 2018). The boosting paradigm works by combining several weak learners, which are typically decision trees, and subsequently using each learner to minimize the errors of the previous learner while maximizing the effectiveness of the previous learner as well. As a result, gradient boosting algorithms tend to perform significantly better than bagging algorithms (Osman et al., 2021; Susheelamma & Ravikumar, 2019). In the XGBoost algorithm, the gradient boosting technique is further enhanced by optimizations in both the software and hardware aspects of the algorithm. For instance, the algorithm has inbuilt parallelization capabilities that make computations significantly fast and cost-effective. Another intrinsic characteristic that makes the algorithm powerful is the tree-pruning technique it utilizes. Unlike typical gradient boosting algorithms which use greedy stopping criteria for their tree splitting, the XGBoost algorithm uses a depth-first approach that prunes trees in a backward approach. This technique is achieved by specifying the *max_depth* parameter of the trees to be used in the algorithm. XGBoost is also widely used due to its additional algorithmic capabilities that are lacking in typical ensemble algorithms. These include aspects like regularization through L1 and L2 techniques which effectively mitigate overfitting, cross-validation capabilities in each iteration of the algorithm as well as sparsity awareness that makes the algorithm ideal for handling missing values (Chen et al., 2018). The XGBoost algorithm was used to predict ground-water levels in Selangor, Malaysia. The direction of the trees is negative to the gradient primarily because the objective of the algorithm is to minimize the overall residual error. This representation was obtained from an investigation of ground-water level predictions in Malaysia using XGBoost (Osman et al., 2021).

Figure 2.25 illustrates the operation principle of gradient boosting algorithms like XGBoost.

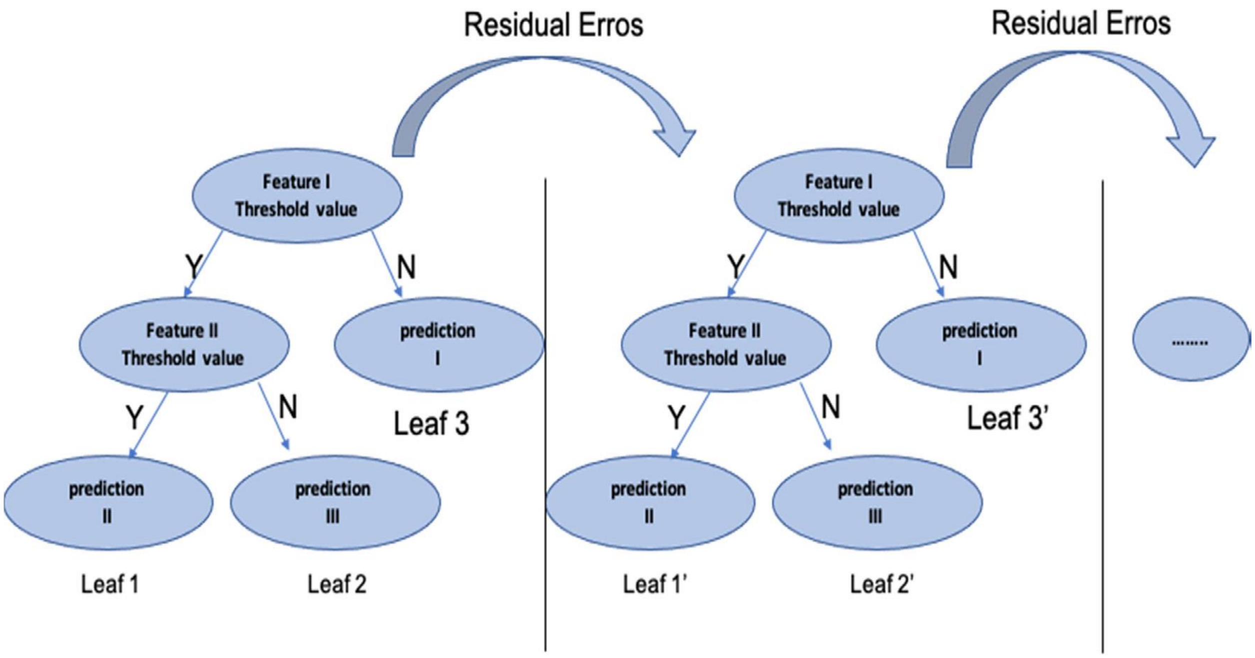


Figure 2.25: XGBoost Algorithm used to Predict ground-water Levels

2.8 Research Gap Analysis

In this section, empirical studies were analyzed to investigate the existing EDM models used in identifying and promoting SRL in online learning environments. The objective was to establish whether the EDM models provide both SRL measurement and promotion concurrently and the approach used to provide the interventions.

Various empirical studies have been carried out in the field of SRL since the emergence of EDM. In the studies, researchers employ various EDM techniques and algorithms to determine learner behaviors, prediction of academic performance, and provision of SRL interventions to enhance learners' skills (Liu et al., 2017; Pardo et al., 2018). The studies indicate that both data collected using self-report tools and data collected from online learning environments is to predict and improve learner performance and retention rates. The studies also investigate the relationship between learner behaviors and performance. For instance in their study, Doleck et al. (2015) investigate how learner patterns influence learner performance or failure while Cavalcanti et al. (2018) developed a model to predict the performance of students based on self-regulated learning skills using EDM methods to identify SRL indicators from datasets collected from an online learning system. The prediction is based on SRL indicators and learners' behaviors, motivation, and application of cognitive abilities.

There has been notable progress made in two areas that provide a foundation for this study. The first is the advancement of the SRL measurement tools used in measuring SRL in online learning environments. The second is the emergence and adoption of EDM in measuring SRL. These areas give insights into how data mining techniques can be integrated into e-learning systems to provide SRL measurement and promotion.

Although some studies have deployed EDM techniques on SRL measurements, there is a need for empirical research to determine the effect of EDM interventions that are based on SRL evaluation from the LMS data, especially on how the interventions stimulate learners to engage in metacognitive activities that can eventually lead to improved academic performance. Additionally, research on provisions of interventions through EDM on Learning Management systems is scanty. Literature indicates that there are two categories of interventions provided to students who study in e-learning environments; the first is through interactive feedback through visualized dashboards and the second is offering metacognitive and behavioral prompts and hints that aim at engaging students to enhance their SRL capability. The use of visualization through dashboards allows learners to see their learning behaviors without stimulating the development of SRL skills. The effect of feedback or dashboard is passive and is described as one-way feedback. According to Viberg et al. (2020), visualized learner-centered dashboards do not enhance knowledge transfer. Moreover, web-based dashboards are limited in terms of access and usage by students both in informal and formal learning environments.

Moreover, we must address how the interventions are delivered to the students in online learning environments and their impact on the students' SRL skills. Literature reveals that instructors first receive visualized reports on dashboards for their exploration and synthesis to understand how individual learners behave. This in turn helps the instructors know what kind of support and guidance to offer the students based on the visualized reports (Araka et al., 2020). The limitations with the models of delivering SRL interventions are that; first, is the reliance on the limited number of instructors to receive and interpret the reports before offering interventions to a large number of students (Nussbaumer, Hillemann, & Albert, 2015). Secondly, the visualized dashboards may require experts who already know how to use the systems and hence may be a limitation to instructors who have little knowledge about the use of computers. While such models are believed to provide information to instructors so that they can make decisions based on facts, instructor judgments may differ hence students may end up receiving varied interventions based on each instructor's interpretation (Baker, 2016).

In summary, the models for SRL promotion can be classified into two (Araka et al., 2020). First, models that extend the decision-making to instructors to enable them to offer students data-driven interventions. This class of models relies on the instructor's knowledge and reports from data analysis on dashboards. Secondly, models that offer metacognitive interventions that trigger students' thought processes and hence have them take charge of their learning. As indicated in this subsection, most of the current studies focus on models that extend decision-making to instructors. However, given the large number of students who enroll in online or blended classes, it may be time-consuming and burdensome for instructors to provide personalized interventions to every student. Studies, therefore, need to focus on bridging this gap and design and examine models that will afford individualized SRL interventions such as hints and metacognitive prompts, empirical evidence on how this can be established is scanty (Lodge et al., 2019).

Despite the progressive research on the use of EDM and its potential in enhancing SRL, a research gap exists. First, the existing EDM models do not provide metacognitive interventions that stimulate students thought processes hence have them develop SRL skills as they learn. Second, there is lack of robust EDM model that perform EDM achieve self-regulated learning amongst online learners. Additionally, the existing models do not link SRL in educational psychology and SRL on LMS. There is need therefore, for an EDM model that contains algorithms derived from learning theories as applied in SRL to provide real-time interventions to online learners.

2.9 Research Conceptual Framework

The research conceptual framework for this study has been applied to carry out the true experiment. The framework, therefore, describes the relationship between the approach for delivering interventions and their effect on students' SRL strategies. The provision of SRL interventions to foster learners to actively participate in learning in terms of prompts is also considered. The approaches used to provide interventions to students; instructor-led and EDM-based interventions are also included. The variables and their relationship are shown in Figure 2.26.

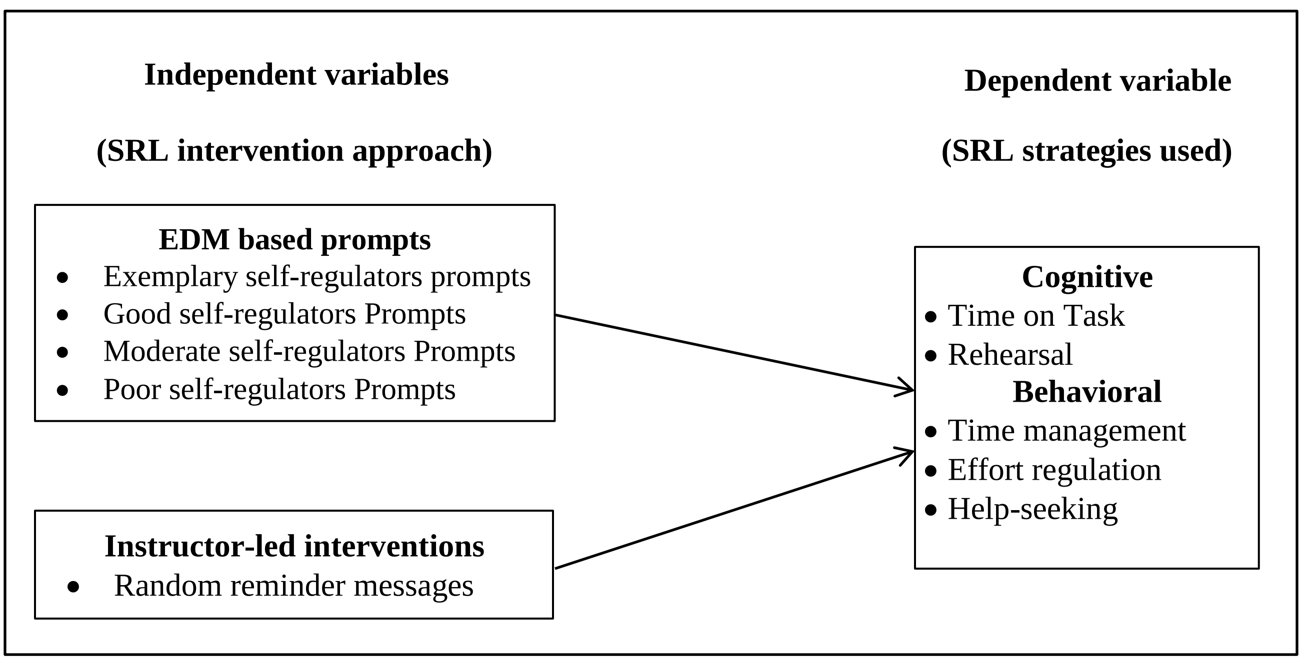


Figure 2.26: The Hypothesized Direction of Influence between Study Variables

The independent variables for this study were derived from the approach used to provide SRL interventions. The two methods that formed the independent variables include: instructor-based interventions and EDM-based interventions guided on the level of self-regulatory skills of the students based on the clusters students are laced. Instructor-led interventions were sent out randomly by the course instructor to the students. For EDM-based interventions, the level of SRL skills for each student was used as an indicator and was based on the profiling of learners into 4 clusters; poor self-regulators (cluster 0), moderate self-regulators (cluster 1), good self-regulators (cluster 2), and exemplary self-regulators (cluster 3). Each of the clusters was assigned numbers 0, 1, 2, and 3 respectively. The LMS log data was used to cluster learners based on their self-regulatory skills levels.

The independent variable, the SRL interventions approach, comprised of; (i) EDM-based interventions, and (ii) Instructor-led interventions. The dependent variable was derived from the SRL strategy being promoted. The SRL strategy has two variables; (i) cognitive strategies and (ii) behavioral strategies. The cognitive strategy has two sub-variables; time on task and rehearsal. The behavioral strategy has three sub-variables; time management, effort regulation, and help-seeking. The variable, method of SRL interventions, is expected to influence the SRL strategies for the students.

2.10 Operationalization of variables

The identification of self-regulation levels for students was based on the analysis of LMS log data which was used to profile learners into four clusters using the clustering algorithm. The classification algorithm was then used to place students into the four clusters identified by the clustering model; poor self-regulators, moderate self-regulators, good self-regulators, and exemplary self-regulators. The four clusters were based on the frequency of SRL activities that students applied to their learning. For example, when students post and reply on discussion forums there could be an indication that students are seeking help from the student peers. This in turn was measured in terms of the number of forum posts created by students and the number of replies to posts by other students. The indicators for SRL strategies were defined from LMS log data. The identification of SRL strategies for learners required that inferences be made from the activities that learners engaged in during the learning process. While the psychological variables such as motivation, cognitive and metacognitive are unobservable features, learner activities are observable and therefore were used to provide inferences that were used to identify patterns that helped to understand the unobservable elements of SRL.

Table 2.13 shows the cluster characteristics that were associated with the four levels of self-regulated learning.

Table 2.13: Cluster Characteristics Associated with Levels of Self-Regulated Learning

Cluster Name	Cluster Characteristics	Reference
Poor Self-Regulators (Cluster 0)	The student logins on to LMS rarely spends few hours on LMS per week, has low or no number of content views, completes quizzes before deadlines, has irregular study intervals (login and online stay period), does not create forum posts and or reply to posts created by other students.	(Ainscough et al., 2019; Çebi & Güyer, 2020; Gašević et al., 2017; Kim et al., 2018; Peach et al., 2019)
Moderate Self-Regulators (Cluster 1)	The student access LMS often, spends moderate hours on LMS per week, has a moderate number of content views, completes quizzes before deadlines, has regular study intervals (login and online stay period), is averagely involved in creating discussion forum posts, and replies to posts created by other students.	(Ainscough et al., 2019; Çebi & Güyer, 2020; Gašević et al., 2017; Kim et al., 2018; Peach et al., 2019)
Good Self-Regulators (Cluster 2)	The student access LMS often, spends more hours on LMS per week, has a high number of content views, completes quizzes before deadlines, has regular study intervals (login and online stay period), is actively involved in creating forum posts, and replies to posts created by other students.	(Ainscough et al., 2019; Çebi & Güyer, 2020; Gašević et al., 2017; Kim et al., 2018; Peach et al., 2019)
Exemplary Self-Regulators (Cluster 3)	The student logins on to LMS very often, spends more hours on LMS per week, has the highest number of content views, completes quizzes before deadlines, has regular study intervals (login and online stay period), actively involved in creating forum posts and replies to posts created by other students.	(Ainscough et al., 2019; Çebi & Güyer, 2020; Gašević et al., 2017; Kim et al., 2018; Peach et al., 2019)

Table 2.14 shows the summary of the SRL strategies, constructs, and learner tasks measured from within an LMS and forms the basis for cluster characteristics.

Table 2.14: Summary of the SRL strategies, constructs and Indicators

SRL Strategy	SRL Constructs	LMS Indicators
Cognition	Time of Task	▪ Time spent on viewing content/resources ▪ Time spent on each online activity (chat/forum/quiz)
	Rehearsal	▪ Content views ▪ Resources views ▪ Course/module information views
Behavior	Time Management	▪ Login (regularity) intervals ▪ Login frequency ▪ Total login time
	Effort Regulation	▪ Number of learning activities completed on time ▪ Number of new activities started by the student (chat) ▪ Number of replies to forums created by others
	Help-seeking	▪ Number of forums posts created ▪ Number of views on forum posts created by others ▪ Numbers messages sent to the course instructor

2.11 Summary

Literature indicates that SRL is an important aspect of learning especially for online learners. The main aim of having e-learning students who can self-regulate is because of limited support and guidance compared to traditional face-to-face classroom setup. E-learning students manage their schedules and decide what time to study and how long to engage in learning. The success and effectiveness of e-learning, therefore, depend largely on the students' ability to play an active role in their learning process. Students who possess high self-regulatory skills perform better compared to those who possess low levels of self-regulatory skills. Additionally, students' academic achievement can be determined and predicted by measuring the level of engagement in an LMS (Barnard et al., 2010; Broadbent & Poon, 2015; Cavalcanti et al., 2018; Cicchinelli et al., 2018; Hashemyolia et al., 2014). This makes Self-Regulated learning an important aspect, especially for students who are studying through online environments such as LMS and have limited interaction with lecturers. The measurement and promotion of SRL on e-learning platforms have advanced from the use of self-report methods which relied on learner's perception of self-regulatory skills and the use of offline data to online measures such as the use of learner logs from e-learning environments to the current opportunities offered by artificial intelligence such as the use of educational data mining and learner analytics. According to Baker (2016), there should be a shift from offering instructor-centered and human-centered student interventions to interventions that are based on educational data insights. Learning management systems provide a rich set of features that if utilized can allow e-learning students to develop and implement self-regulatory skills in learning. From the existing literature, it was noted that most institutions of higher learning utilize LMS for disseminating learning materials while paying less attention to how learners engage in online activities which are supposed to stimulate the growth of SRL skills. Dabbagh and Kitsantas (2013) records that utilization of LMS features and functionalities are significantly related to the growth of self-regulatory skills such as self-evaluation, task strategies, self-monitoring, time management, goal setting, and help-seeking. LMSs, therefore, provides a great opportunity to promote SRL skills for learners (Vovides et al., 2007).

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the research philosophy adopted by the researcher, the research methodology applied to conduct the study, the unit of analysis, sampled population, data collection, and data analysis methods. This chapter is organized as follows:

Section 3.2 describes the research philosophy adopted by the researcher. Section 3.3 gives an overview of the research methodology. Section 3.4 describes the quantitative research methodology used to carry out the pre-study that was used to establish the current status on how LMS features are utilized by students to promote SRL. The study also formed the foundation for establishing the system requirements analysis for prototype development. Section 3.5 describes the integrated design process that was used for system development. Section 3.6 discusses the true experiment research methodology that was used to evaluate the EDM model.

3.2 Research Philosophy

The researcher in this study holds a pragmatic worldview. The pragmatic research approach offers the freedom to choose research methods, tools, and techniques that solves the research problem. Pragmatists are concerned with the solution that works to solve the problem and that mixed approaches can be used to understand and solve the problem (Creswell, 2003).

3.3 Overview of the Research Methodology

With this pragmatic worldview, the current research used a mixed-methods approach where four methodologies were applied; qualitative and quantitative research designs, Integrated Design Process (IDP) for system development, and true experimental research design for model evaluation.

The research process was carried out in three phases which include problem identification, solution design, and evaluation as described in Figure 3.1 (Offermann et al., 2009).

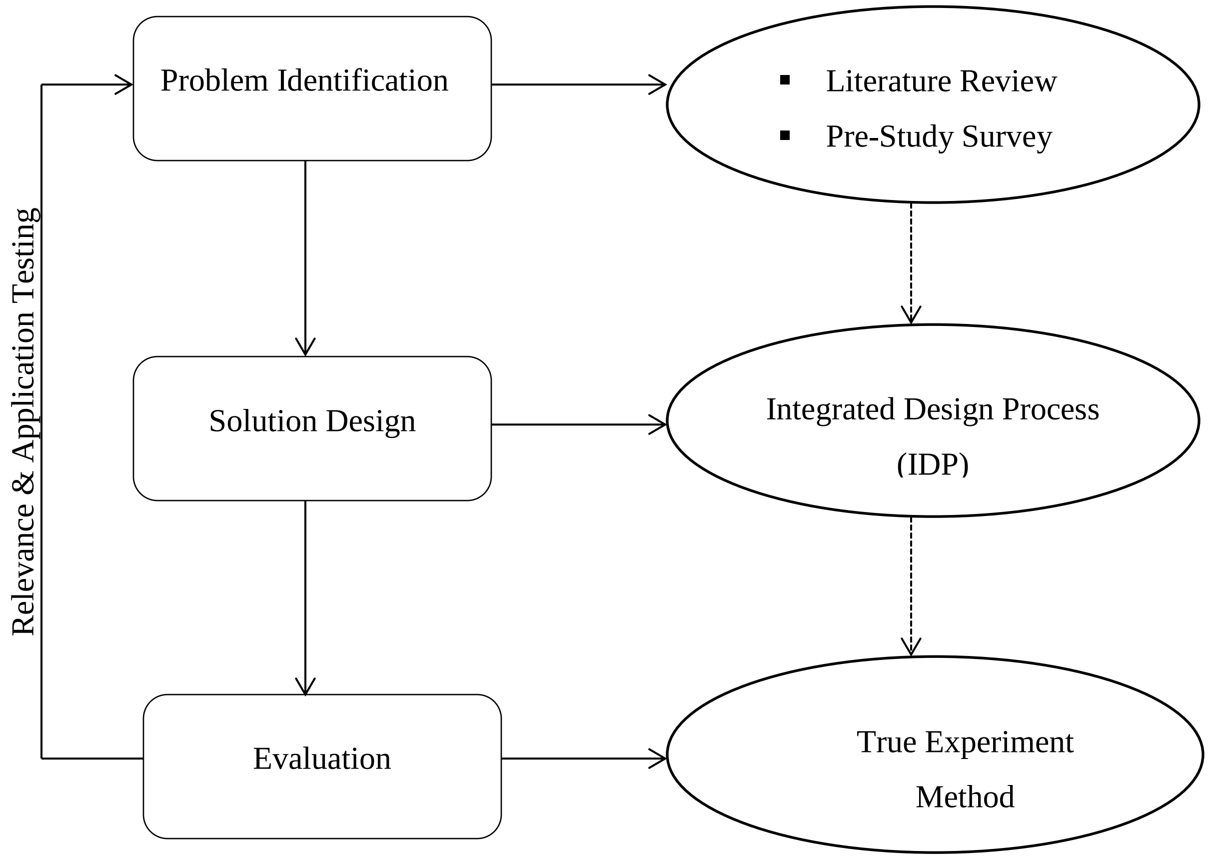


Figure 3.1: Research Design Process

The first phase of the research involved the use of both qualitative and quantitative methods for problem identification. First, qualitative research was used to conduct document analysis to explore the current methods being used to measure and promote SRL in online learning environments. Secondly, a quantitative research method was used to carry out a pre-study survey using a structured questionnaire to establish the current status on how LMS features are utilized in teaching and learning in higher institutions of learning in Kenya.

The second phase of the research involved the design and development of an intelligent model that incorporated EDM techniques as informed by the pre-study survey and relevant theories from literature (Offermann et al., 2009). The IDP was adopted for system design. The methodology consists of four phases of system development; needs analysis, conceptual design, development, and evaluation (Nam & Smith-Jackson, 2007).

The third phase of the research involved the evaluation of the EDM model using the true experiment research methodology. The methodology was used to examine the effect of EDM-based interventions on students' SRL strategies. Two systems namely an experimental system and a control system were used to provide an online course to two randomly assigned groups of students. After the experiment, a post-study on the experimental group was carried out through a structured survey and a semi-structured interview. The structured survey was conducted to investigate the students' perception of the usefulness of the EDM model in promoting SRL using the interventions provided based on their cognitive and behavioral skills as inferred from LMS logs. The semi-structured interview was carried out to establish students' perception of the benefits of the intervention messages that were provided during the course through the LMS.

Figure 3.2 illustrates the mapping of the research objectives and questions to research methodologies used in the study.

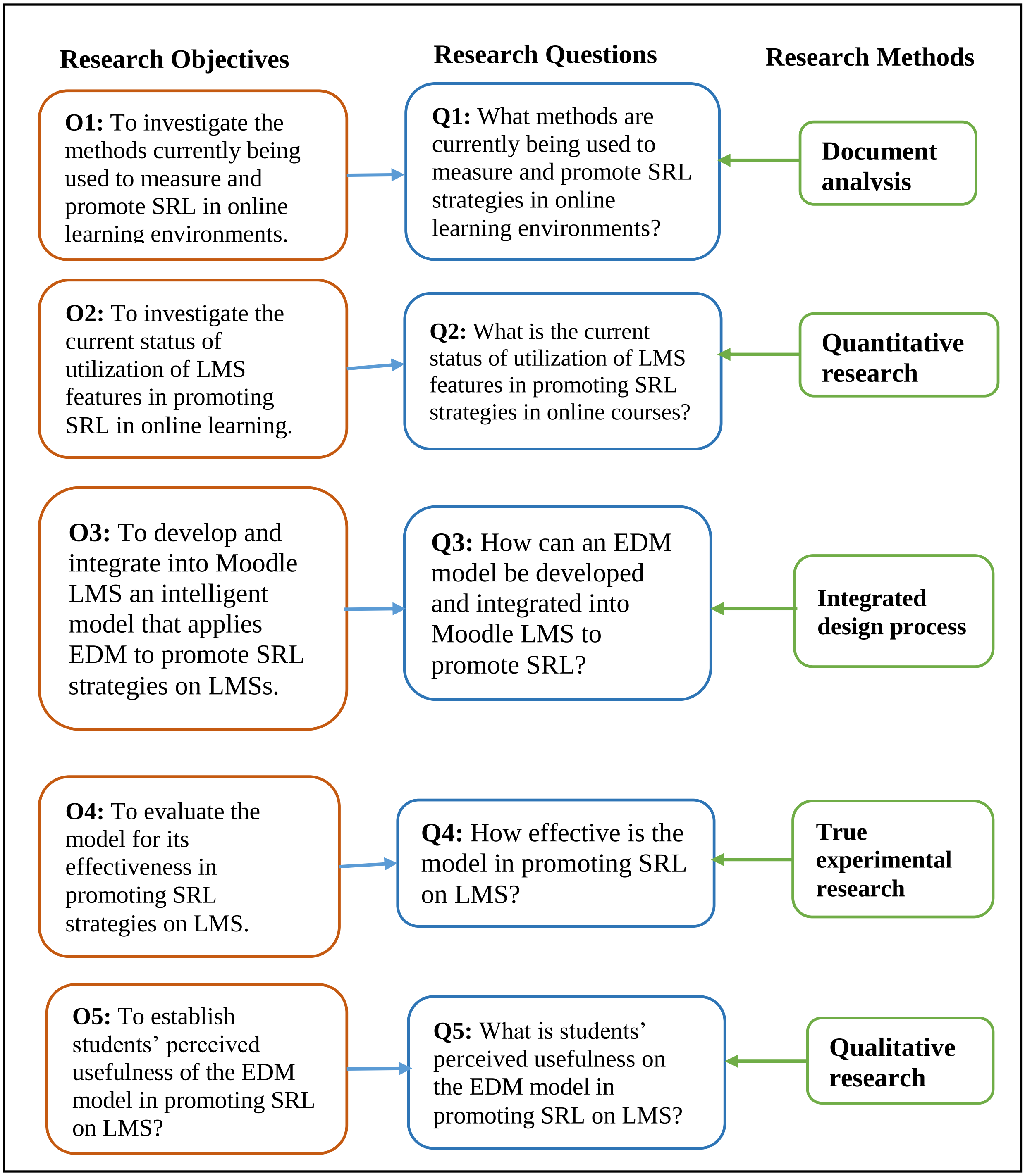


Figure 3.2: Mapping of Research Objectives and Research Questions to Research Methods

3.4 Pre-study Research Methodology

This section describes the descriptive research methodology that was used to establish the perception of university students pursuing their courses through the open and distance e-learning (ODEL) where learners enroll for fully online courses, the data collection instrument that was used to collect data from the study participants, and how the collected data was analyzed. The pre-study was conducted to examine the current status of utilization of LMS features in promoting SRL in higher institutions of learning. The study moreover sought to investigate the students' perception of the utilization of LMS features in promoting them to utilize self-regulatory skills in online learning. In view of this, the pre-study was guided by the following questions:

1. What is the extent to which LMS features are utilized to support SRL during an online course?
2. What is the extent to which SRL strategies such as time management, self-evaluation, self-monitoring, help-seeking, effort management, and organization are utilized by students during an online course on LMS?
3. What are students' perceived challenges associated with the utilization of LMS features in promoting SRL?

3.4.1 Target Population

The structured survey involved University students who were pursuing online courses through the ODEL mode. Five universities, including Maseno University, Mount Kenya University, Kenyatta University, Cooperative University, and Kenya College of Accountancy University were purposively sampled. The purposive sampling technique was preferred as it allows efficiency and robustness as the researchers need to only establish what is being investigated and identify the right people and population who will provide the information required for the study (Dolores, 2007). The selected universities offer open and distance learning programs where students pursue online courses with minimal interaction with lecturers in face-to-face mode. An online invitation was sent to students who had studied online courses for at least two semesters through LMS. A total of 700 emails for students were obtained and received voluntary invites to take part in the study via emails that contained a link to the questionnaire provided through an online survey tool. Out of the 700 students who were invited, 495 responded. This was a 71% response rate.

3.4.2 Data Collection Tool

The descriptive research was carried out using a structured questionnaire that contained 28 items identified in the literature on student engagement in online learning environments and self-regulated learning (see appendix IV).

To ensure the reliability and validity of the tool, the questionnaire was first sent to LMS and SRL experts for review. The experts' responses were used to refine the instrument. Qualitative data was also collected through open-ended questions to establish students' perceptions and challenges they experience in online learning.

The summary of the items and information that was obtained from the questionnaire is presented in Table 3.1.

Table 3.1 Summary of the Items and Information Obtained from Survey Instrument

Questionnaire Item	Type of Question/Scale	Information collected
1 to 7	Multiple choice	Demographic characteristics
8 and 9	Multiple choice	The type of LMS features and tools used for communication between students and instructors
10	Likert scale	How often was the communication between students and course instructors
11	Likert scale	How often do the students utilize various LMS features such as forums, quizzes, chats, blogs, and wikis in an online course
12 -15	Multiple choice	Utilization of LMS features and functionalities in promoting SRL
16 and 20	Multiple choice	The LMS features and tools used by instructors to monitor students' learning process
17	Multiple choice	The factors that hinder students from taking an active role in online learning
18	Likert scale	How often do the students receive feedback aimed at improving student's learning habits
19	Multiple choice	The source of feedback received by students
21	Multiple choice	Areas that students receive little support from course instructors
22	Multiple choice	Self-Regulated Learning strategies the students utilize most in an online course
23	Likert scale	How often do the students receive feedback on performance in submitted assignments and quizzes
24	Likert scale	Level of satisfaction the students received on various issues in regards to individualized support
25	Multiple choice	The SRL strategies ever utilized by students through the use of various LMS features and tools
26	Multiple choice	The various approaches that students feel will be most appropriate in promoting various SRL strategies
27	Open-ended	The additional features students think the e-learning system should have for supporting individualized learning
28	Open-ended	The students experience in online courses in terms of the support they receive from instructors

3.4.3 Data Analysis

Descriptive analysis was carried out on quantitative data while content analysis was performed on qualitative data. Content analysis on qualitative data was carried out to understand the students' perceived challenges in the utilization of LMS features in promoting SRL in online learning environments. The data was analyzed using the R studio statistical tool.

3.5 System Development Methodology

The Integrated Design Process (IDP) methodology was used to develop the experimental system that provided the intelligent promotion of SRL. The IDP which was proposed by Nam and Smith-Jackson (2007), emphasizes the design of a dynamic instructional system that suits learners' interaction while being guided by the cognitive and constructivism theories of learning (Oboko, 2012). The IDP design approach also focuses on the solution design process that strives to fulfill system requirements from the theory building and pre-study survey (Offermann et al., 2009). The IDP consists of four phases of system development; needs analysis, conceptual design, prototype development, and system testing as illustrated in Figure 3.3 (Nam & Smith-Jackson, 2007).

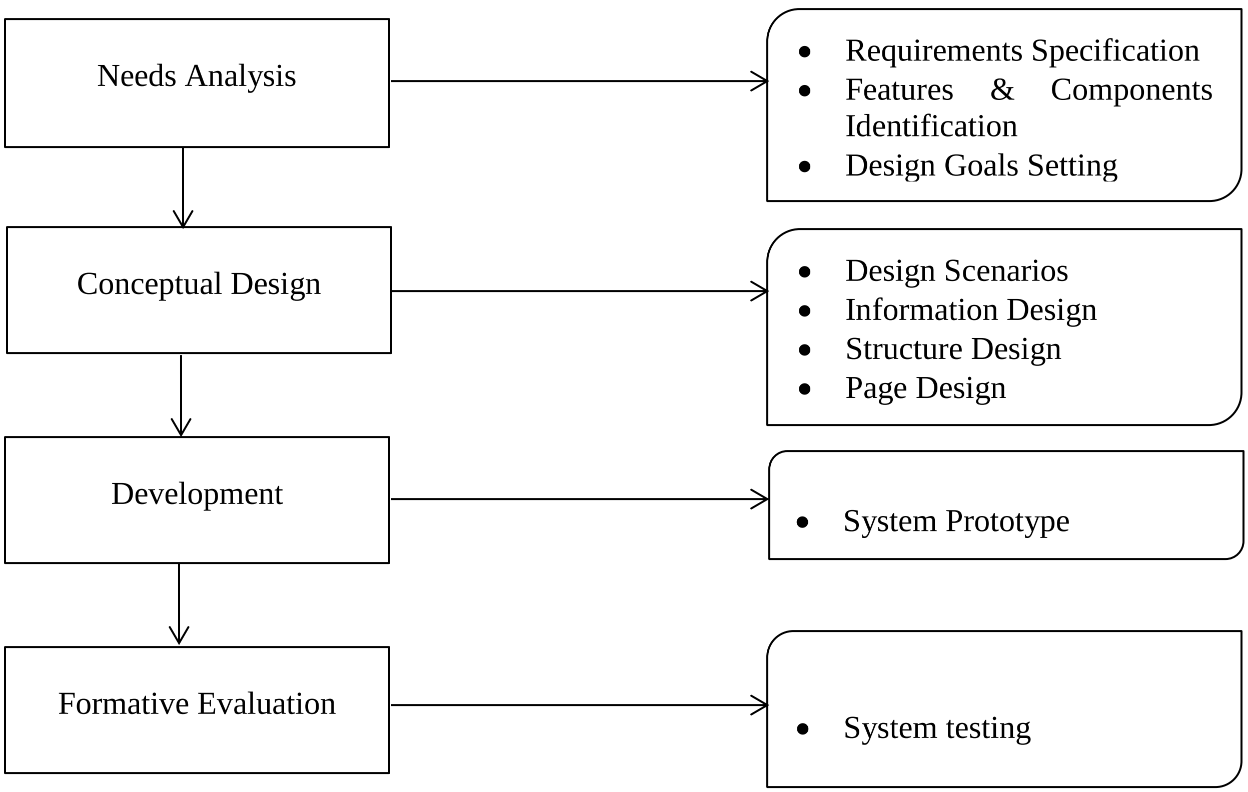


Figure 3.3: Integrated Design Process

The needs analysis phase involves specifying user and system requirements (while considering the target group) necessary to build the system. The features and components of the system are also identified at this phase. The second phase involves putting the concepts together on how the system will reflect the user's activities. The concepts then produce a conceptual design of the system. In the third phase, the development phase, a prototype is constructed from the conceptual design. The fourth and last phase involves carrying out a formative evaluation that involved testing the workability of the prototype. The testing aims to discover and implement any area that might have been overlooked in the previous phases.

3.5.1 Needs Analysis

The needs analysis phase involved specifying learner and system requirements that guided the development and implementation of the experimental learning management system. The needs analysis phase comprised of three processes that were followed in collecting and analyzing the system features; requirements specification, features and components identification, design goals setting, and identification of intervention messages The needs analysis phase also involved the identification of learner requirements that informed the design of the experimental system. The identification of learner requirements was obtained using the pre-study tool (see appendix IV).

a. Requirements Specification

The requirements specification process involved the identification of features that captured system specifications that support the provision of targeted interventions to learners based on their SRL levels. The information was gathered using the pre-study tool (see appendix IV) and literature review. The students' feedback was collected through the pre-study on students' perception of LMS features in promoting SRL skills, the online learning experiences, and their expectations towards enhancing current LMS that supports them in enhancing their SRL skills.

b. Features and Components Identification

The features and components identification process involved the identification of the system features that support the learners' requirements specification that was identified during the requirements specification process.

c. Design Goals Setting

The design goals setting involved the determination of system design goals that formed the foundation for evaluating the experimental system. The input to the design process included the system and learner requirements and the features that were identified from the preceding process; requirements specification and features and components identification processes.

Table 3.2 presents the general design goals that were identified for system development.

Table 3.2: General Design Goals Setting

Design Goal	Description
Impact	Strengthen learners' motivation and self-regulatory behaviors
Simplicity	Easy navigation to students' dashboard
Clear mapping	Provide students with a summary of achieved activities and motivate them to engage in what is pending and needs to be done
Informative communication	Allow learners to receive and follow the suggested learning activities through metacognitive prompts
Dynamic	The intelligent module places learners dynamically in the SRL group they belong to and provide targeted prompts and alerts to the student. If there is engagement improvement re-allocate the student to the respective SRL level automatically

Based on the inputs from the previous processes and the general design goals, the following design objectives were set for system development.

1. To provide a user interface that allows learners to receive individualized support through SRL interventions
2. To provide a system that enables learners to play an active role in managing their learning processes
3. To allow for students to receive real-time metacognitive prompts to enhance their learning processes
4. To provide a system that presents a summary of learning activities to students via a visualized dashboard.
5. To design a system that profiles learners into clusters based on the similarity of their online learning behaviors

The design goal-setting process also involved identifying the specific design goals for an online learning environment that supports the theory of constructivism and the promotion of self-regulatory skills to learners.

Table 3.3 presents a summary of the design goals for an LMS that supports constructivism and the provision of prompts identified from existing literature.

Table 3.3: Summary of the Design Goals for LMS that supports SRL Promotion

Design Goal	Description	Reference
Self-Monitoring	The learners should be able to reflect and take action for learning outcomes.	(Araka et al., 2021; Nam & Smith-Jackson, 2007)
Time management	The learner should be able to spend more time on the course	(Pérez-Sanagustín et al., 2020)
Self-Evaluation	The learners should be able to pause and think about their learning processes before proceeding	(Nam & Smith-Jackson, 2007; Rowe & Rafferty, 2013; Wong et al., 2018)
Help-seeking	The learners can seek help and consult with their peers in the course	(Lee et al., 2019)
Increased Learner engagement	Collect data on how students' engagement behaviors, analyze the data into information to be used to guide a student in improving their engagements in learner activities	(Araka et al., 2021; Nam & Smith-Jackson, 2007)
Learner support feedback	The system to provide individualized feedback messages to students based on their engagement behavioral patterns	(Araka et al., 2021; Nam & Smith-Jackson, 2007)

d. Identification and Design of Intervention Messages (Prompts)

The intervention messages (prompts) identified in literature were classified according to the framework proposed by (Wirth, 2009). The framework defines cognitive, metacognitive, and behavioral activities under which prompts can be presented to learners. The SRL strategies that were being promoted via the prompts were also defined. The conditions that triggered the presentation of the prompts to the students were based on the type of learning activities students engaged in. The conditioning enhanced the provision of targeted interventions as they were based on real-time analysis and placed students in respective clusters. The framework also provides behavioral and cognitive prompts to learners based on their previous learning activities as inferred from the LMS log data. For higher effectiveness and immediate benefit to the students, the prompts were provided immediately after they logged into the LMS and during the learning episode depending on how triggers were conditioned and guided students' behavior and cognition (Sedrakyan et al., 2018).

3.5.2 Conceptual Design

The conceptual design process involved converting the system requirements identified in the needs analysis phase into a design that illustrated the functionality of the experimental system. The process also involved conceptualizing how to integrate the user requirements into Moodle LMS. The conceptual design involved gathering the system design concepts from the system specification and the features and components process from the needs analysis phase and producing a high-level conceptual design of the experimental system. The conceptual design involved the following process that was followed in converting system requirements into the conceptual design:

a. Design scenarios
b. Instructional module design
c. System architecture design
d. Conceptual Interface design

a. Design Scenarios

The design scenarios process involved the identification and conversion of requirements specifications into tasks that could be carried out by learners and enabled the EDM module to capture the tasks through the associated components. The tasks that were captured allowed learners to interact with LMS, peers, and course instructors.

Sample design scenarios and learning tasks carried out by students are presented in Table 3.4.

Table 3.4: Sample of the Design Scenarios and learning tasks carried out by students

Task	Object	Physical action
Login to LMS	Course main page	Logins into the system by entering access credentials
Checks out on LMS announcements or receive promptings directing the learner to engage in specific learning activities	EDM visualization dashboard	Finds out the interventions messages received via the link to the plugin
Carry out pending learning activities	Course navigator	Select the suggested learning activities
Get involved in the learning activities	Course topic links or learning activities links	Clicks on learning activities scheduled

b. Instructional Module Design

This process involved the development of the instructional module for the course that was selected for evaluation of the EDM model in Moodle LMS. The main tasks included two tasks; designing the structure and developing the course content. The output from the instructional module design process was the course outline, course content, materials, and the relevant learning activities for each topic covered. Course instructors developed the draft of the course module and identified the learning activities that were integrated within the course modules. The activities included discussion forums, quizzes, chats, blogs, and journals that were integrated into the course content. The module development process involved three Data Science and Machine Learning lecturers who developed the module. A few experts were invited to give their views and also rated the module if it was standard to be offered to University students. Finally, educational experts in online instructional development vetted the course. The observations and feedback from the subject experts; one subject expert and an instructional design expert were used to correct and refine the course module that was uploaded into Moodle LMS.

Table 3.5 shows the validation feedback received from experts who were involved in the review of the module for the data science with python course.

Table 3.5: Validation Feedback from Experts on Instructional Module

Expert	Area of expertise	Feedback received
Instructional Design expert	Senior Lecturer at Kenyatta University, Kenya. Experienced in the design and delivery of blended learning materials; Content Development for blended delivery and online learning	• Replace *lesson objectives* with *learning outcomes*. An outcome is what the student can do by the end of a lesson. An objective is what we want to achieve as teachers by the end of the lesson • From lessons 6-9, reflect on the student's activities once more and clarify what you want students to do? Under individual contribution, specify what you want students to participate in. Do not generalize. • Consider getting supplementary resources from the *IBM Digital - Nation Africa* on Artificial Intelligence and Machine learning • Develop lessons 9 and 10 completely with learning outcomes and instructions on what you want the students to do.
Subject expert	Data Science lecturer at the School of Mathematics and Computer Science, University of Wolverhampton, UK	• The structure is fine, and suitable for online delivery, with of course support of an LMS • The module is largely Machine Learning and not so much Data Science. However, this does not affect the change of content to be delivered to the students.

After revising the module based on the feedback from the experts, the final module was developed for uploading into Moodle LMS.

c. System Architecture Design

The system architecture design process involved translating the conceptual design into an architectural design that described the components of the experimental system concerning the integration of the EDM plugin that was used promotes cognitive and behavioral skills. The process also described how the design elements for the system were organized. The process involved two tasks; visualization of the conceptual design and converting the conceptual design into the system architecture using the standard format for visualization (Nam & Smith-Jackson, 2007; Oboko, 2012).

d. Conceptual Interface Design

The interface was designed as per the guidelines from the findings from the pre-study survey guided by the needs analysis and conceptual design phases. The plugin interface is comprised of a student dashboard and instructor/administrator dashboard. First, the students' interface was designed to present to the students the summary of the learning activities they engaged in and the list of the interventions messages that have been sent to them on Moodle LMS interface through pup-ups. Secondly, the instructor interface was designed to allow instructors to view the number of students clustered, and the distribution of students amongst clusters while indicating the summary of learning activities carried out by each student.

Figure 3.4 shows the students' interface.

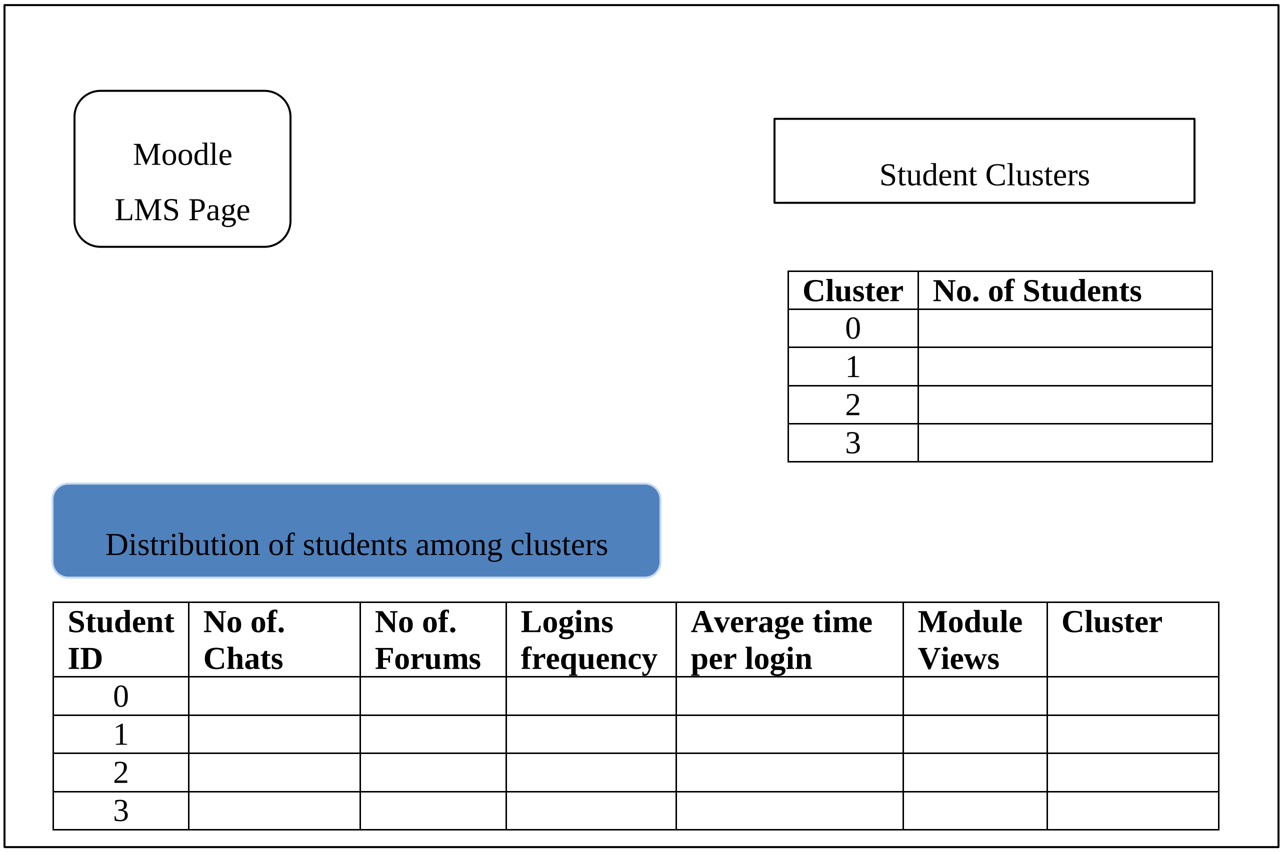


Figure 3.4: Instructor Interface Design

The students' interface was designed to display progress or summary of students’ learning activities and prompts container which contained the most recent messages sent to the student and the number of messages sent. The message prompts were presented to learners on the Moodle LMS pages.

3.5.3 System Development

The development process involved converting the conceptual design into a working prototype. The system design concepts were translated into an experimental working prototype. The input to the development phase included the requirements specification, system features and components, instructional module content, conceptual design, interface design, and system architecture design.

3.5.4 System Testing

System testing was carried out to evaluate the workability and functionality of the experimental system. The testing phase involved evaluating the working prototype to identify areas for improvement and revisions. The system testing involved the enrollment of students and inviting them to study a short course on the Basics of Python for four weeks. A sample of 54 students was purposively sampled and registered into the experimental system. The students were taken through a short course that lasted for one month. The students were invited to access the materials uploaded content on Python Basics. The data collected was used for clustering and classification of the student into similar clusters and targeted prompts provided. The system was evaluated based on two metrics; (a) classification of students into clusters, (b) display of the analytics dashboard that displayed a summary of learning activities for each student, and (c) the provision of interventions to the students.

3.6 Experimental Research Methodology

3.6.1 Introduction

This section describes the experimental research methodology that was used to evaluate the effectiveness of the EDM model for promoting SRL in Moodle LMS. The true experimental research design methodology was used to examine the effect of EDM-based interventions on students' SRL strategies (cognitive and behavioral skills). Two systems; an experimental system and a control system were used to provide an online course to two randomly assigned groups of students. After the experiment, three datasets were obtained;

(a) LMS log data obtained from the experimental system and control system
(b) Quantitative data was obtained using the MSLQ questionnaire to collect time-series data from students who participated in the courses before and after the course
(c) Quantitative and qualitative data were obtained from a survey using a structured questionnaire to investigate students' perception of the effectiveness of the EDM plugin and from a semi-structured interview that was carried out to qualitatively identify students' perceived benefits of interventions provided by the EDM plugin and establish areas of improvement and additional features.

This study investigated the educational data mining techniques that can be applied to educational data generated when studying online to provide targeted interventions to support the individual needs of the learners. Providing individualized support for online learners in higher institutions of learning is a major challenge as large numbers of students are enrolled in online courses. Exploring LMS log data using EDM techniques offers great insights into understanding students learning behaviors, therefore, opening an opportunity for instructors and educators to establish the type and nature and interventions for various groups of students. Additionally, the LMS features which are designed to place learners at the center of their learning as active agents are underutilized learners. Literature reveals that there is a lack of evidence on the effect of providing targeted interventions based on learner behaviors as captured by the trace data. In view of this, there was a need to explore the data generated when students are engaged in online learning activities, analyze the data, and cluster learners into similar groups to allow for the provision of targeted interventions for each group of learners. Moreover, there was a need to investigate how SRL interventions can be offered based on the trace data from LMS and the effects of the interventions on supporting the online learning process (Schumacher & Ifenthaler, 2020). To evaluate the effect of promoting students' SRL strategies by offering interventions based on students' online learning behavioral patterns the experiment in this study were carried out to establish answers to the following research questions;

a) How effective is the model in promoting SRL on LMS?
b) What is students' perceived usefulness of the EDM model in promoting SRL on LMS?

Research indicates that when students learn on their own, guided by constructivism theory (von Glasersfeld, 2014), they perform better and are more likely to be satisfied than those who do not. However, students underutilize the LMS features that enable them to engage more and interact more with learning resources and learning activities. Learning interventions in form of prompts and feedback messages motivate learners to utilize LMS features and therefore reinforce their SRL skills during the learning process. Moreover, the SRL interventions scaffold and place learners at the center of learning, therefore, improving their levels of engagement and interaction with learning materials and learning activities. Supporting online learners and motivating them to interact with LMS features stimulates the development and use of SRL skills in the learning process. Students with high levels of SRL perform better in terms of academic grades compared to those with low SRL skills. Based on the above research questions, the following hypotheses were formulated for testing;

- N**ull Hypothesis**: There is no significant difference between SRL strategies utilization by students in EDM-based intervention and instructor-based intervention groups.
- **Alternative Hypothesis**: There is a significant difference between SRL strategies utilization by students in EDM-based intervention and instructor-based intervention groups.

The rest of this chapter describes the study population, research design, data collection, and data analysis techniques that were used to test the hypothesis and address the research questions.

3.6.2 Study Population

The study population was students from institutions of higher learning in Kenya. First, convenience sampling was used to identify accessible sample in Nairobi County. Second, purposive sampling was used to identify three universities (Kenyatta University, The Co-operative University of Kenya, and the University of Nairobi) in the county. Purposive sampling was used since it does not require consideration of underlying theories of SRL. The method is also efficient and robust as the researchers need to only establish what is being investigated and identify participants with the characteristics being considered in the study (Dolores, 2007). Students (who were pursuing bachelor's degree courses related to Computer Science, Statistics and Mathematics, and Information Technology) from the sampled universities were voluntarily invited to register and participate in the data science with python course via Google form. The data science course was preferred due to the interest students have developed in machine learning as the core course in the artificial intelligence field. The data science course is not offered in the current curriculum across the universities, yet is a preferred professional course for students pursuing computer science, IT and statistics related programs. The number of students who registered for the course was 696 and those who completed the course were 231.

3.6.3 Research Design

This study adopted a true experimental research design which involved the use of a non-equivalent pretest-posttest control group design. The true experiment involved randomly assigning participants into two groups; the experimental group that received automated interventions on the LMS and the control group where students received randomly sent out messages by the instructor. All students in the two groups received the same learning materials, learning activities, and the LMS administrative support. Table 3.6 shows the treatments received by each group.

Table 3.6: Experimental and Control GroupTreatments

Group	Pretest	Treatment	Posttest
Experimental	E_1	T_1 = EDM-based interventions (based on the level of their SRL)	E_2
Control	C_1	T_2 = Instructor-led interventions (random)	C_2

Source: (Oboko, 2012)

The research process took 13 weeks that included three-phase. First, the participants were randomly assigned into two groups and a pretest was carried out. This phase took one week. The second phase was the learning phase where the students were subjected to online learning on Moodle LMS for ten weeks. Finally, phase three which took two weeks involved assessing the students, post-test, and post-study survey.

The experimental research design process is summarized in Figure 3.5.

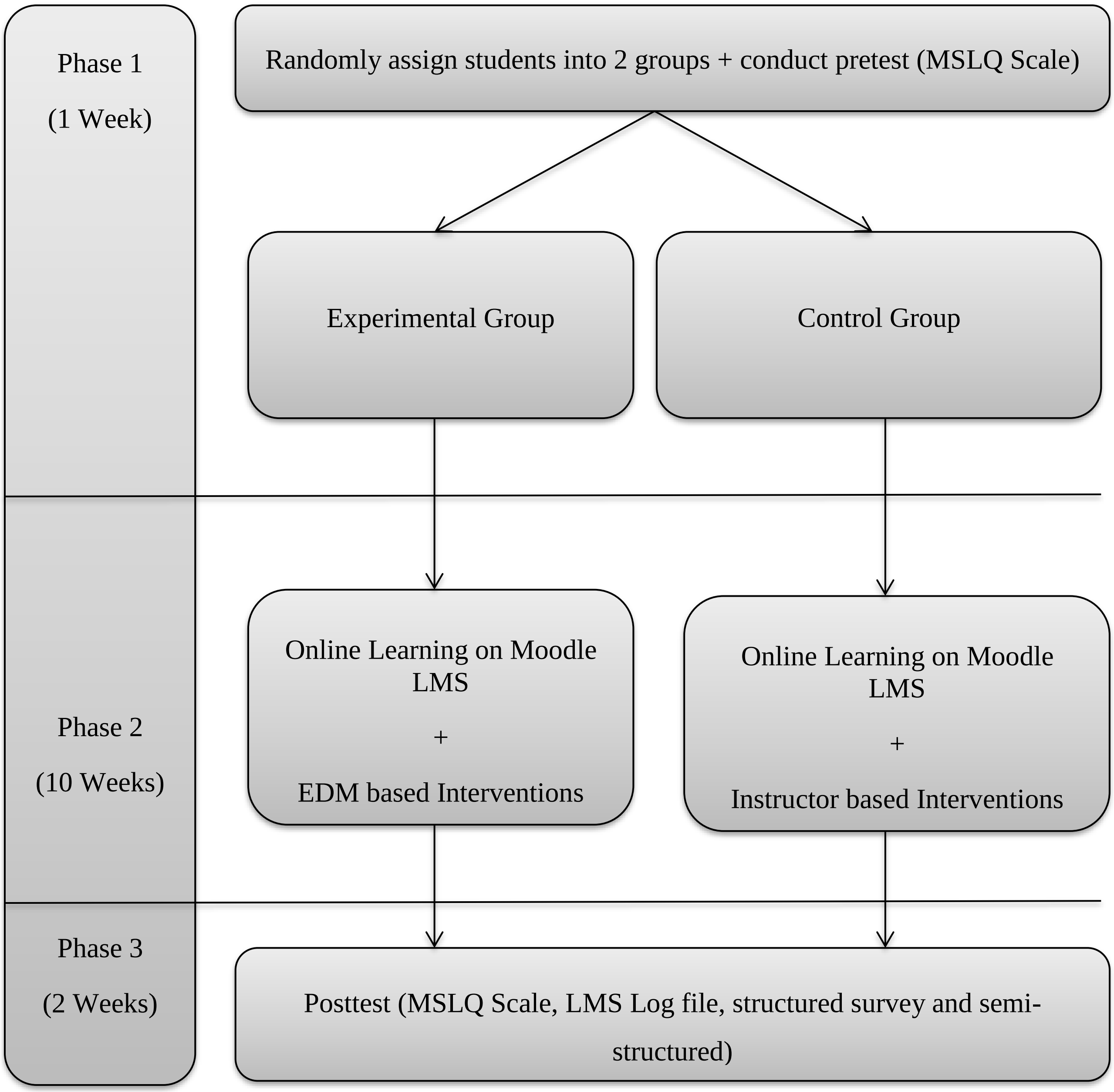


Figure 3.5: Experimental Design Process

The three phases were conducted as follows;

a) Phase 1: Registration & Pretest

Students were invited to register to participate in the online course. After registration, the students were randomly assigned into two groups; the experimental group and the control group. The students were then invited to respond to the MLSQ scale which was used to collect pre-test data. After the pre-test, the students were separately enrolled in experimental and control groups

b) Phase 2: Actual Learning

Phase two of the research design involved the actual learning process. The experimental group was enrolled in the course through the experimental system with the EDM plugin and the control group enrolled in the control system without the EDM plugin.

c) Phase 3: Assessment, Posttest, and Post-studies

The students from the two groups took a summative assessment test provided through the Moodle LMS. To do the assessment, students were first invited to respond to the MLSQ scale which was used to administer a post-test. Additionally, students in the experimental group participated in a semi-structured post-study survey and a semi-structured interview that was conducted to investigate the students' perception of the usefulness of the EDM plugin in promoting SRL strategies in online learning.

3.6.4 Data Sources and Instruments

This section described the sources of data and the data collection instruments that were used in the experimental process. First, the motivated strategies for learning questionnaire (MLSQ) scale was used for the pretest and posttest. Time series data from the Moodle LMS and EDM plugin was collected for analysis to explore how students employ SRL skills in learning. A structured questionnaire was used to collect students' perceptions of EDM interventions provided to the experimental group. Finally, a semi-structured interview was carried out to collect qualitative data.

a. Data Sources

This section describes the sources of data that were explored and collected for analysis in this research; Moodle LMS log files, EDM Plugin log files, and Moodle LMS grade books. The Moodle LMS log files from both experimental and control systems that comprised the traces that were indicators of engagement levels and indicators of SRL strategies were collected after the experiments. The log data captured students' interactions with learning content, videos, and learning activities such as chats, forums, assignments, and quizzes. The logs from Moodle LMS were obtained and used to examine the online learning behavioral patterns and SRL skills for both the experimental and control groups in the study. The EDM plugin log files collected from the experimental systems' database that was integrated into Moodle's core database were collected. The data comprised of the students that students were placed progressively and the messages they received throughout the learning process. Table 3.7 describes the data sources used for this study.

Table 3.7: Description of Data Sources

Data Source	Description
Moodle LMS log files	The log files contain students' interaction with the online course for both experimental and control groups. The logs contained the interactions students had carried out and recorded on Moodle LMS
EDM Plugin log files	The log files contained students' interaction with the EDM plugin as captured by Moodle LMS core database and the integrated plugin. The logs also contained students' summaries of activities they engaged in and the visualizations on the plugin dashboard. The log file also captured students clusters over the learning period
Quantitative and Qualitative	Quantitative data was obtained using the MSLQ tool to collect time-series data from students who participated in the courses before and after the course. Quantitative and qualitative data were obtained from a survey using a structured questionnaire that investigated students' perception of the effectiveness of the EDM plugin and from a semi-structured interview that was carried out to qualitatively identify students' perceived benefits of interventions provided by the EDM plugin and establish areas of improvement and additional features.

b. Pretest and Posttest Instruments

The Motivated Strategies for Learning Questionnaire (MSLQ) was used to measure students' SRL skills before and after the online data science with python course. The MSLQ is an SRL measurement scale that was developed by Pintrich et al. (1991) to measure the motivation and learning strategy levels of students and has subsequently been validated as a reliable and effective SRL measurement tool (Khampirat, 2021). The scale contains two subscales motivation and learning skills each with 31 and 50 items respectively. The rating for the items is based on a 7-point scale where 1 = not at all true of me to 7 = very true of me. The summary of the MSLQ items and their subscales are described in Table 3.8.

Table 3.8 SRL Strategies captured in the Motivated Strategies for Learning Questionnaire

Scale	SRL Strategies
Motivation	• Intrinsic Goal Orientation • Extrinsic Goal Orientation • Task Value • Control of Learning Beliefs • Self-Efficacy for Learning and Performance • Test Anxiety
Learning strategies	• Rehearsal • Elaboration • Organization • Critical Thinking • Metacognitive Self-Regulation • Time and Study Environment Management • Effort Regulation • Peer Learning • Help-Seeking

Source: (Duncan et al., 2005)

This study considered the subscales that contained the learning strategies under investigation. The strategies include resource management strategies: effort regulation, peer learning, help-seeking, and time and study environment management. For metacognitive/cognitive strategies: rehearsal, elaboration, organization, and critical thinking. The scale adapted for this study contained 32 items involving the control of learning beliefs (4), self-efficacy for learning and performance (8), rehearsal (4), time and study environment (8), effort regulation (4), and help-seeking (4). For more details on the specific items captured refer to Appendix V and VI.

Table 3.9 presents the summary of the items that were measured during the pretest and posttest (also refer to appendix V and VI respectively).

Table 3.9: Summary of SRL strategies contained in the MSLQ used for pretest and posttest

Scale	Sub factors	Item numbers from the original scale
Motivation	Control of Learning Beliefs (4)	2, 9, 18, 25
	Self-Efficacy for Learning and Performance (8)	5, 6, 12, 15, 20, 21, 29, 31
Learning strategies	Rehearsal (4)	39, 46, 59, 72
	Time and Study Environment Management (8)	35, 43, 52, 65, 70, 73, 77, 80
	Effort Regulation (4)	37, 48, 60, 74
	Help-Seeking (4)	40, 58, 68, 75

c. Post Study Survey and Interview Instruments

This section describes the tools that were used to conduct post studies that involved students from the experimental group. First, a structured survey was conducted with 141 students from the experimental system. Second, a semi-structured interview that involved 20 students who volunteered to participate was conducted.

i. Post-study Structured Questionnaire

The structured questionnaire was developed to explore the students' perception of the EDM tool used to promote SRL in LMS. A structured survey was conducted on the students in the experimental group to investigate the students' perception of the usefulness of the EDM model in promoting SRL using the interventions provided based on their cognitive and behavioral skills as inferred from LMS logs.

Table 3.10 presents the summary of the items that were captured on the questionnaire.

Table 3.10: Summary of the Item and Information obtained from the Post Study Survey

Item	Type of Item/Scale	Information Collected
1-3	Multiple choice	Demographic features
4	Multiple choice/Likert scale	Tools students use to communicate amongst themselves and instructors in class
5	Multiple choice	Students' previous experience in blended or online learning
6	Multiple choice	Students' previous experience in Data Science and Machine Learning course
7	Multiple choice	Analytical and statistical tools students knew before starting the Data Science with Python course
8	Multiple choice	The factors and issues that hindered students from actively involved in online learning during the course
9	Likert scale	Students' online learning experience during the Data Science with Python course
10	Likert scale	How LMS helped students utilize SRL strategies during the Data Science with Python Course
11	Likert scale	Students' perception of the usefulness of the support they received in form of intervention messages and email notifications
12	Open-ended	The best online learning experience students had during the Data Science with Python course
13	Open-ended	The worst online learning experience students had during the Data Science with Python course
14	Open-ended	The key areas or features students suggest to be included in the interventions provided by LMS to enhance online learning

ii. Post Study Semi-Structured Interview

A semi-structured interview was carried out with students who had completed the online course using the experimental system that provided intervention messages to learners. The purpose of the interview was to establish students' perception of the benefits of the intervention messages that were provided during the course through the LMS. The participants were voluntarily invited to participate in the interview that involved the following four open-ended questions;

1. Based on your experience, how was the course?
2. Could you briefly explain how the intervention messages from the LMS supported you during the course?
3. What other intervention features do you think can help support you to be in control of your learning?
4. Do you think there is a better way you can be getting the notifications from the Learning Management System?

The interview was carried out using the Google Meet link and then transcribed using the otter.ai application.

The threats to the internal validity of the data collection instruments and experiment designs and the measures that were put in place to mitigate the threats are presented in Table 3.11.

Table 3.11: Internal Validity Threats and Mitigation Measures

Threat	Mitigation measures
Independence of observations	Participants were randomly assigned into groups and a pre-test was carried out to ensure uniformity across participant characteristics and external bias. Participants in each group did not know that the other group existed as they were separately registered and enrolled in the online course at individual levels.
Validity of the data collection instruments	Expert validation was applied for each instrument used.
Internal Consistency of the data collection instruments	The Cronbach's alpha values were computed and the values obtained were within the acceptable range of 0.65-0.75.
LMS facilitation and external interference	• All the participants in each group separately received instructions on what was expected during the online learning course and were given guidelines on the Moodle LMS and its features and usability • Same learning materials and learning activities were used for the two groups and were subject to the same learning period. • The researcher in this study was not involved in the facilitation of the online course • Two separate instructors facilitated the online learning and none of them knew which system they were using (i.e. experimental or control) • Two experts in the SRL prompts were involved in the design of the prompt messages that were used in the experimental group during the online course

3.6.5 Procedure

After the registration for study participants was closed, the data was obtained and randomly assigned into two groups. First, the data were randomly shuffled and split into two halves. One half was assigned group 1 and the other half group 2. Secondly, a column was introduced indicating the group of the students. Group 1 was assigned to the experimental group and group 2 was assigned to the control group. After randomly assigning participants into experimental and control groups, students from each group were separately invited to respond to the MLSQ tool for the pretest. The registered students were randomly assigned into two groups. The experimental group was enrolled in the experimental system where the EDM plugin had been installed and the control group enrolled in the control system without the EDM plugin. The experiment was carried out in 12 weeks where a data science with python course was provided to the students using two systems; the experimental system and the control system.

The pre-test was carried out after the students were randomly assigned into two groups and were used to test for any significant differences between the two groups before engaging in the online course. After the online course which took ten weeks, the post-test was carried out using the MLSQ tool to measure students' SRL skills and used to investigate the SRL levels for the experimental group and control groups. The post-test was used to test for any significant difference between the groups who participated in the study.

There were no prompts that were provided externally or outside the learning environment by the instructor for each group. The instructor only debriefed all the students once a week. The control group only received instructional support messages on Moodle LMS via the announcement component. No instructional support or prompts were provided to the control. A participation certificate of completion was issued to all the participants who completed the course. The EDM plugin provided students with SRL interventions inform of prompts and feedback based on learner engagement behavior by grouping students into four distinct clusters. The clusters reflected the different levels of behavioral engagement in the learning as defined by the EDM model.

3.6.6 Learning Environment

The Moodle LMS was used to deliver the data science with python course to the students. The course comprised 10 lessons. Each lesson comprised of learning materials and learning activities that included; notes, videos, assignments, chats, discussion forums, external URLs, and quizzes.

All the learning materials were provided through the LMS and students were not allowed to download the notes or any other learning materials to an external environment. This guaranteed all the learning activities carried out by students created adequate trace data for analysis to understand learning behaviors (Cerezo et al., 2020). All the students received the same learning materials and the same number of online learning activities, all initiated by the course instructor. The Moodle LMS captured students' learning activities and interactions with learning materials for the both control and experimental group. The control group used LMS without the plugin installed and received instructor-based support messages which were presented to them via the LMS's announcement component. The experimental group used LMS with the EDM plugin installed. The learners received EDM-based intervention messages sent out to them in real-time within Moodle course as pop-up messages and also email notifications.

3.6.7 Data Analysis

This section describes the methods that were used to analyze the data that was obtained in this study. The study consists of quantitative data collected from the actual learning management system and pre-study survey, pretest, post-test, and post-study survey and interview. A mixed-method approach was used to analyze the data that was obtained in this research.

First, the Shapiro-Wilk test was done to test form normality in the distribution of pretest and posttest data. A p-value of less than 0.05 was obtained for all the items indicating the normal distribution of data. Inferential and descriptive statistics were applied to quantitative data obtained from pretest, posttest, and post-study surveys and time-series data obtained from Moodle LMS. Content analysis and inductive thematic analysis were used to analyze qualitative data from the post-study survey and semi-structured interview. An independent t-test was used to examine the effectiveness of the EDM tool in promoting SRL on LMS.

3.6.8 Summary

This chapter describes a mixed methods research approach that was carried out in this study that involved three phases. First, descriptive research using a structured survey was used to conduct a pre-study that formed the foundation for system development. Secondly, the integrated design process was used to develop the system that was used for model validation. The EDM model was implemented using a clustering algorithm that was used to identify four clusters of learners based on historical data obtained from Moodle LMS course. A classification algorithm was also developed and integrated into the model. The classification algorithm identifies students (new instances) who access the LMS and places them in their respective clusters to receive targeted interventions. The provision of EDM-based interventions was implemented through a pop-up widget within Moodle LMS pages and weekly email notifications that targeted students who rarely accessed LMS for the online course. The system was tested using students invited into a short course who were not involved in the main experiments described in this research. The system was evaluated using a true experimental research design methodology. The true experimental research design methodology involved examining the effect of EDM-based interventions on students' self-regulatory skills. Students from higher institutions of learning, who formed the unit of analysis in this research, were randomized into two groups; experimental and control groups. After random assignment, a pretest was administered to test for homogeneity between the groups in terms of the SRL skills before offering the treatments during the experiment. The experimental group was enrolled in an LMS with the EDM plugin installed while the control group was enrolled in an LMS without the EDM plugin installed. The students in the experimental group received EDM-based interventions during the online course while the control group received instructor-based interventions. After the experiment, a posttest was administered to both experimental and control groups. Moreover, the experimental group participated in two post studies that involved a structured survey and a semi-structured interview to establish students' perceived benefits of the EDM model in promoting self-regulated learning in learning management systems.

CHAPTER 4

RESEARCH FINDINGS AND DISCUSSION

4.1 Introduction

This chapter presents the research findings obtained from the pre-study, system development, and the true experiment and provides a discussion of the findings based on the research questions. The main objective of this study was to investigate the effect of offering EDM interventions to students based on their learning behavior as inferred from the LMS log data. Based on the research questions, the findings were achieved as follows;

The first objective was to investigate the methods being used to measure and promote SRL strategies in online learning environments. To achieve this, a systematic review of literature on the various techniques being used to measure and promote SRL was conducted and the challenges facing these techniques were identified. The study also aimed at identifying the gaps in the current techniques used to promote SRL strategies, especially on LMS. The findings were discussed and presented in sections 2.4 and 2.5.

The second objective was to investigate the current status on the utilization of LMS features in promoting SRL in online learning. To achieve this, a pre-study was carried out using a descriptive research methodology that involved administering a structured questionnaire to students who were pursuing their courses via ODEL mode. The shortcomings of the current LMS features in promoting SRL strategies in online courses were identified. The findings were discussed and presented in section 4.2.

The third objective was to develop and integrate into Moodle LMS an intelligent model that applies EDM to promote SRL strategies on LMS. To achieve this a critical review and an empirical study of the current EDM techniques being used to identify and promote SRL in online learning environments was conducted. The optimal EDM algorithm for identifying SRL profiles was identified. The algorithm was then used to design a prototype, herein referred to as the EDM plugin, which was used to evaluate the effectiveness of the EDM model in promoting SRL on LMS. The plugin was then integrated into Moodle LMS to allow for the evaluation of the EDM model. The development and integration process is discussed and presented in section 4.3.

The fourth objective was to evaluate the effectiveness of the EDM model in promoting SRL strategies on LMS. The evaluation was conducted using the true experimental research methodology as discussed and presented in section 3.4. The findings are presented in section 4.4.

The fifth objective was to establish students' perception of the usefulness of the EDM model in promoting SRL strategies on LMS. To achieve this, post studies involving a structured survey and semi-structured interview were conducted on students who used the experimental system (EDM plugin installed) in their online course. The findings from the post studies are presented in section 4.3.3.

Finally, section 4.5 provides a discussion of the research findings according to the research questions and hypotheses that were formulated and tested.

4.2 Pre-Study Findings

a. Participants' Demographic Information

The demographic characteristics of students involved in the pre-study are summarized in Table 4.1. From the findings, 66% of the respondents were male while 34% were female. The results indicate that 56% of students were undergraduate students while 40% were postgraduate students. The results also indicate that 50% of students engaged fully in online learning while 46% engaged in blended learning where online learning supplemented traditional face-to-face teaching. Most of the students representing 49% spent 1 to 10 hours per week in online learning while those who spend 11-20 and 21-30 hours per week represent 29% and 15% respectively.

Table 4.1 presents a summary of the demographic information of the participants.

Table 4.1 Demographic Features of Sample Population (N=495)

	Demographic feature	N	%
1	**Age (years)**		
	15-25	71	14.34
	26-35	279	56.36
	36-45	113	22.83
	46-55	28	5.66
	Above 56	4	0.81
2	**Gender**		
	Female	167	33.74
	Male	328	66.26
3	**University of Study**		
	KCA University	26	5.25
	Kenyatta University	406	82.02
	Maseno University	5	1.01
	Mount Kenya University	43	8.69
	The Cooperative University of Kenya	15	3.03
4	**Level of study**		
	Certificate	4	0.81
	Diploma	26	5.25
	Postgraduate	197	39.80
	Undergraduate	268	54.14
5	**Mode of learning**		
	Blended learning	230	46.46
	Full-Time student	9	1.82
	Fully Online	248	50.10
	Part-time or Evening	8	1.62
6	**Hours spent online per week**		
	1-10 hours	244	49.29
	11-20 hours	144	29.09
	21-30 hours	72	14.55
	31-40 hours	20	4.04
	41 or more	15	3.03

b. Students’ Responses on Utilization of LMS Features in Promoting SRL

From Table 4.2, it can be observed that quizzes, chats, discussion forums, and messages were regularly used by students while wikis, workshops, and blogs were least utilized amongst the students.

Table 4.2 Utilization of Various LMS Features by Students (N=495)

LMS feature	Never (%)	Often (%)	Rarely (%)	Sometimes (%)	Very Often (%)
Forums	2.83	36.57	12.32	25.45	22.83
Chats	1.62	40.00	7.47	24.44	26.46
Wikis	40.40	6.87	20.81	28.08	3.84
Emails	4.44	30.91	12.12	32.12	20.40
Messages	6.46	28.08	12.32	31.52	21.62
Quizzes	5.45	38.18	5.66	17.98	32.73
Blogs	38.38	9.49	24.24	25.45	2.42
Workshops	38.79	10.10	22.63	25.45	3.03

The present study captured the purpose for which various LMS features were utilized amongst online students. Table 4.3 tabulates the reasons for which chats, forums, and messages were utilized. As indicated, most students used chats to seek help. For example, 32% of students indicated that they use chats to seek help from other students while 30% used chats to seek help from the course instructors. The students who used chats to meet course goals represent 27%. The students used discussion forums to meet course goals and seek help from course instructors as represented by 34% and 28% respectively while 42% of the students use messages to seek help from fellow students and 27% used messages in seeking help from instructors.

Table 4.3 The Purpose for which LMS Features were Used by Students (N=495)

Purpose	Chat (%)	Forum (%)	Message (%)
Meet course goals	26.94	33.52	17.86
Meet strategy goals	7.76	10.27	6.41
Seek help from course instructors	30.32	27.67	26.56
Seek help from other students	32.02	23.82	41.53
I am not aware of the tool	1.27	1.43	2.44
I have never used the tool	1.69	3.28	5.19

c. Students' Responses on Utilization of SRL Strategies

In Table 4.4, we demonstrate how various LMS features were utilized to promote the use of SRL strategies amongst students. As indicated 32% of students utilized quizzes in achieving self-evaluations while 24% indicated that they use quizzes for self-monitoring. Forums and chats were also mostly utilized by students to achieve various SRL skills. For example, 43% of students used chats to seek help from instructors and students. The discussion forums were used to utilize almost all the SRL strategies provided in the study. The results show that emails, wikis, and workshops were the least used to achieve the various SRL strategies.

Table 4.4 How Various SRL Strategies were Utilized by Students using LMS Features (N=495)

SRL strategy	Chat %	Email %	Forum %	Quiz %	Wiki %	Workshop %	None of the tools %
Time management	28.11	8.43	30.52	18.74	3.48	1.74	8.97
Self-monitoring	24.82	5.92	29.44	23.81	4.18	2.45	9.38
Self-evaluation	23.1	4.45	27.69	31.99	4.59	2.30	5.88
Help-seeking	42.9	12.81	25.91	6.55	4.60	2.09	5.15
Effort management	25.85	9.45	28.66	16.69	5.47	3.10	10.78
Organization	24.89	8.35	29.21	13.71	6.41	2.68	14.75

The results on how students utilized SRL strategies in online courses are presented in Table 4.5. The results indicate that 23% of students have utilized a self-evaluation strategy, and time management 19%, while 14% utilize help-seeking strategy and 13% utilize effort management in online learning. As indicated, 14 % of students said that they were not aware of any SRL strategies while 3% said that they have never utilized any of the SRL strategies.

Table 4.5 How Students Utilized Various SRL Strategies in Online Courses (N=495)

SRL Strategy	N	%
Self-evaluation through the use of quizzes and self-assessment tools on LMS	260	22.97
Time management by using LMS features to plan time use e.g. frequent logins and when to respond to various activities shared by instructors	216	19.08
Help-seeking by seeking assistance from peers or instructors through question and answer	154	13.60
Not aware of any SRL strategies	154	13.60
Effort management by actively participating in online learning by having frequent logins and reminders to view learning materials. Know the effort to put towards achieving your course goals	145	12.81
Organization by scheduling learning activities for the day or week	108	9.54
Self-monitoring through the use of tracking or journaling tools to monitor learning progress	58	5.13
None of the above	37	3.27

d. Students' Perceived Challenges in Utilizing LMS Features in Promoting SRL

The students perceived challenges associated with the use of LMS features in promoting SRL are presented in this section. Table 4.6 presents the challenges that hinder the students from playing an active role during an online course.

From the results, it can be observed that 28% of students indicated that "*lack of adequate internet*" is a major hindrance to engaging in online study. "*Lack of interaction with course instructors*", and "*lack of individualized feedback on students' learning habits*" were the main factors that hinder students from actively being involved in online learning at 19% and 15% respectively. Other factors include "*Lack of instructor guidance*", and "*Lack of peer interaction*" at 14% and 11% respectively.

Table 4.6 Factors that Hinder Students from Actively Participating in Online Learning (N=495)

Hindrance to active participation in online learning	n	%
Lack of adequate internet	234	28.47
Lack of interaction with course instructors	154	18.73
Lack of individualized feedback on learning habits	127	15.45
Lack of instructor guidance	113	13.75
Lack of peer interaction	90	10.95
Other hindrances	62	7.54
Lack of adequate learning	42	5.11

The study also captured the areas students perceived there is little support from course instructors as presented in Table 4.7. Most of the students indicated a lack of real-time and individualized feedback at 22% and 20%. From the results, 15 % of students indicated that they rarely receive guided learning while 12% said that there is little support in the provision of prompts that guide them on their study habits.

Table 4.7 Areas Students Perceive there is Little Support from Course Instructors (N=495)

Area with little support	n	%
Real-time feedback	207	22.43
Individualized feedback	186	20.15
Guided learning	136	14.73
Prompts guiding you on your study habits	112	12.13
Instructional help	92	9.97
Study hints	92	9.97
Provision of learning materials	83	8.99
Others	15	1.63

e. Students' Experiences on Support from Course Instructors

From an open-ended question, students were asked to give their experiences on the kind of support they received from instructors during online courses. Table 4.8 presents some of the highly cited experiences by students.

Table 4.8 Students' Experiences on Support Received from Instructors

Area of support	Sample Responses
Instructor feedback/interaction	"Instructors are not usually available for support and offering individualized assistance", "most lectures do not respond to online queries", "online course is very difficult as there is no interaction with instructors like full-time", "instructors not active, most lecturers take too long before responding to student's chats", "I have not received any support from lecturers", "good but online interactions need more effort", "need more interaction."
Online monitoring, guidance, and support	"lack engagement", "there is not much online guidance", "very minimal support making the course hard because of limited time", "the support was negligible", "the support is very minimal", "there is lack of follow-up and prompts feedback from instructors", "I have not received much support and they can improve on that", "online instructors need to be more involved in guiding students."

f. Suggested Additional Features for Promoting Self-Regulated Learning

Through an open-ended question, we sought to establish the additional features students would like to have included in LMS.

Table 4.9 highlights some of the frequently mentioned items.

Table 4.9 Suggested Areas of Improvement in Terms of Additional Features for LMS

Suggested area of improvement	Sample Responses
Instructor Interaction	"Instructor support needs to be boosted", "simple study hints", "one on one communication links", "two-way communication with students and instructors", "real-time response to questions", and "group chats."
Mode of examination administration	"Online exams to make it convenient", "final exams should be done online", "ability to sit for examination online", and "administer exams online."
Lack of adequate internet	"Access to modules without the use of internet or mobile data", "SMS chatting for areas with no internet."
Instructor Feedback	"Real-time response to queries", "more individualized attention", "immediate feedback on assignment and quizzes", "features for prompt feedback"
Integration of social media and other tools	"WhatsApp", "integration with videos", "SMS integration", "Twitter"
Monitoring learner progress	"Progressive alerts as reminders on learner progress", "the University should provide progress bar for monitoring progress with online learners", "tracking device", "prompts on ongoing activities", "real-time notifications", "regular reminders", "well-managed progress bar", "learning tracker to monitor how far one has reached."
Organization and time management	"Time management features", "clear notifications on activities to be done."

g. Students' Feedback on System Intended Functionality

Through an open-ended question, we sought to establish the areas of improvement and the intended functionality of LMS that promotes SRL skills. Table 4.10 presents students' suggested areas of improvement and the intended functionality of LMS that promotes SRL skills.

Table 4.10: Students' Feedback on System Intended Functionality

Suggested Areas of Improvement	Cited Responses
Instructor Interaction	"Instructor support needs to be boosted", "simple study hints", "one on one communication links", "two-way communication with students and instructors", "real-time response to questions", and "group chats."
Instructor Feedback	"Real-time response to queries", "more individualized attention", "immediate feedback on assignment and quizzes", and "features for prompt feedback"
Monitoring learner progress	"Progressive alerts as reminders on learner progress", "the University should provide progress bar for monitoring progress with online learners", "tracking device", "prompts on ongoing activities", "real-time notifications", "regular reminders", "well-managed progress bar", and "learning tracker to monitor how far one has reached."
Organization and time management	"Time management features", and "clear notifications on activities to be done."

h. Pre-Study Findings that Influenced System Development

The pre-study findings also helped identify the features that had a direct influence on the design of the experimental system. The pre-study carried out on students from selected universities in Kenya who also served as intended users were used to gather feedback on whether the current LMS supported them in improving SRL skills. The survey also helped identify the learning activities mostly performed by students and the expected needs of the students. Additionally, the pre-study findings that had a direct influence were also considered. The findings from the pre-study survey helped capture requirements specification and system design. Table 4.11 presents the pre-study findings and how they influenced the system development process.

Table 4.11: The Pre-Study Findings that Influenced System Development

Result	Influence on System Development
Lack of individualized feedback	Design an algorithm that assigns students into groups based on their learning behaviors. The similar groups of learners are then used to offer timely feedback based on their behaviors as captured by learning activities engaged on
Lack of instructional support	Include metacognitive prompts to guide students in their study
Minimal involvement in learning activities based on chats, discussion forums, and quizzes	Collect log files on the use of these features and use the analysis to support learners through prompts

4.3 System Development Findings

This section describes the findings from the system development process from the identification of features and components to the system integration into live Moodle LMS. The features and the corresponding components were identified by reviewing the students' feedback from the pre-study tool (see appendix IV). The consolidated list of all the features and their corresponding components that were captured in the experimental system are presented in Table 4.12.

Table 4.12: Features and Components of the Experimental System

Feature	Relationship with LMS	Component
Individualized Support	Provide individualized support through the use of prompts to the learner	Messaging mechanism that points the learner to what they are supposed to do or what they have missed doing.
Active Learner	Allow and motivate students to be in charge of their learning processes	A flexible navigation feature that allows learners to be directed to the learning activities to engage in to make them play an active role during learning.
Real-time Feedback	Provide metacognitive feedback to the learner to allow them to stop learning and reflect on their learning behaviors before proceeding	A student model that reinforces the use of real-time feedback to help place learners as active players in the learning process
Learner profiling	Group learners into clusters of similar learning behaviors to enable the provision of targeted interventions through metacognitive prompts	• An intelligent model that profiles learners into clusters based on their SRL levels. • Fetch and provide interventions to targeted groups
SRL clustering	Provide a mechanism to dynamically allocate students into respective SRL levels based on learning activities completed and hence behavior patterns	Dynamic SRL clustering model that identifies the change in students' behavioral patterns and places them in the corresponding class

SRL interventions manager	Provide metacognitive prompts	• Visualized dashboard which indicates the current position of the learner, what the learner has achieved and what is pending • Prompt each learner according to their learning behaviors
Dynamic allocation of a student into SRL clusters	Allow dynamic placement of students into relevant SRL levels depending on the learning behavior to allow receive targeted interventions	Clustering algorithm
Learner Control	Enable learners to actively be involved in learning activities; what to study, what expected learning activities	Metacognitive prompts provider
Help-Seeking	Allow learners to seek help and collaborate with peers using discussion forums and chats	Collaborative activities allocator

4.3.1 Rubric for SRL Interventions for the EDM Model

This section describes the SRL intervention messages (prompts) that were integrated into the EDM model. The prompts were presented to students via the Moodle LMS interface as designed on the plugin. The presentation of prompts was based on students' past learning behavior (Wirth, 2009). Providing the prompts after feedback from the model ensured that the interventions were targeted and individualized for each learner based on their behaviors as captured by Moodle LMS logs. The prompts were associated with the behaviors identified from the LMS indicators and the corresponding SRL strategies. The prompts, which comprised of either question-prompt or question-prompt and recommendation or a combination of both, were timed and presented when the students had access to the LMS hence affording real-time access to feedback on their learning behavior (Wong et al., 2020) and were targeted to promote behavioral and cognitive skills as captured by Pintrich's model (Azevedo, 2009). The prompts were presented to the students via the plugin dashboard inform of pop-ups and associated with recommended learning materials or learning activities. For those students who could not access LMS regularly, email reminders were sent to them twice a week.

The students in each cluster received SRL prompts designed for each cluster's features. The prompts were targeted for each student based on the students' level of engagement with learning activities within the LMS. The condition under which prompts were triggered varied from student to student depending on the frequency of views on resources and content provided via URLs, page views, and students' participation in chats, quizzes, assignments, and forums. The targeted prompts were delivered to the students based on the cluster characteristics. The prompts were classified based on the cluster characteristics which were defined by students' learning behavior and engagement levels. The number of prompts presented to each student was based on the number of times students visited and clicked on the main trigger which was Moodle LMS's *mod_view API* and what the student was able to achieve in each learning activity. The prompts delivered to the students were guided by;

i. Timing and time intervals of the prompts

The intervention messages were presented to learners via a pop-up widget within Moodle LMS during the learning episode. The email reminders were delivered to the students twice a week and were targeted especially those students who were not able to access the LMS with frequencies varying based on the cluster in which the student was placed (Bouchet et al., 2013b).

ii. Content of the Prompts

The content and type of prompts provided to the student were classified either as cognitive or behavioral. Prompts with the combination of both the timing/intervals of presenting prompts and the content of the prompts are more effective in promoting SRL strategies in online learning environments (Engelmann et al., 2021).

From literature, the rubric presented in Table 4.13 was designed and used to identify the SRL intervention messages that were used to promote various SRL strategies in this study.

Table 4.13: Rubric for SRL Intervention Messages for each Cluster

Cluster	SRL Intervention Messages	Corresponding SRL Strategy
Poor Self-Regulators (Cluster 0)	• You have not spent time studying course content and videos, time yourself and spend time studying course materials • Create time to access LMS and participate in the online learning activities such as chats, forums, quizzes, assignments, and reading/listening to the learning resources such as notes and videos	Time on Task
	• According to system records, you have not accessed the online course for nearly a week. Could you be facing any challenges? Consider accessing LMS and seek support from the course instructor via message, chat, or forum. • Did you know that you can pause reading and think about the previous content covered? Try to reflect on what you know with what you are currently reading	Rehearsal
	• Through the system records, we have noted that you have not logged into the online course for nearly a week. Create time and log in and participate in the learning activities • System records indicate that you have spent little or no time on the online course. Spend more time studying course materials and videos to help you understand the course better	Time Management
	• Try to put effort and time into difficult areas of the course to make it easy for you to understand.	Rehearsal
	• You have not attempted any learning activity such as chat, forum, quiz, and assignments. This may lead to poor performance at the end of the course. Create time for online learning!	Help-seeking

	• If you are not doing well in accessing the e-learning system and interacting with peers and instructor, you may consider seeking support from peers and instructor by reading, posting, and replying to chats and forums.	
Moderate Self-Regulators (Cluster 1)	• Your level of interaction with the learning activities resources is low and alarming. Consider creating time for the online learning activities • It is highly suggested that you create time and access LMS and post and reply to other students' forums and chats	Time on Task
	• Remember to pause and remind yourself of what you are studying and try to summarize what you have read in each video and lesson • If you are facing any difficulty in understanding the course content and videos, try re-reading the difficult parts	Rehearsal
	• Your level of interaction with learning resources such as notes and videos is low and alarming. Consider creating time for online learning • It is highly suggested that you create time and access the e-learning system and read and post to other students' forums and chats	Time Management
	• As you go through the course content and videos, take more time to reflect and recall what you are studying • Are there areas in the course in which you are experiencing some difficulty? Try to revisit them and make short notes and you study.	Effort Regulation
	• As you study the online course, try to seek help from your peers and instructor by posting new chats and forums and replying to chats and forums posted by your peers to see if you can improve in your course.	Help-Seeking

Good Self-Regulators (Cluster 2)	• You are making good progress in interacting with learning resources and participating in online learning activities, keep up the spirit • You are spending quality time posting and replying to forums and chats created by the course instructor and fellow students. Good effort!	Time on Task
	• From the system records, your effort in creating and replying to chats and forums is commendable. Good work. Maintain the momentum • You seem to be timing yourself well in engaging yourself with online learning activities such as chats and forums. Maintain the effort and do not be distracted.	Rehearsal
	• Your frequent access and the time you spend on the e-learning system for the online course are high and commendable. Good job!	Time Management
	• As you go through each lesson in the course, take more time to reflect and recall what you are studying • Did you know that putting more effort into study materials and online learning activities enhances your course performance? You are doing well on this. Maintain the efforts.	Effort Regulation
	• Your effort in seeking help and sharing your course knowledge with others by creating and replying to chats and forums is commendable.	Help-Seeking
Exemplary Self-Regulators (Cluster 3)	• You are doing excellent work in the course by engaging and participating in online activities such as chats, and forums. Keep up the spirit	Time on Task
	• You seem to be easily navigating through the online learning activities and learning resources. Excellent!	Rehearsal

	• You are making excellent progress and taking control of your learning. Maintain the rhythm.	
	• Your frequent access to the e-learning system for the Data Science course is commendable. Remember to always take more time on online learning activities and studying course materials such as notes and videos.	Time Management
	• Do not forget to always review what you have covered at the end of each lesson, by making your brief notes. • Try your best to always relate what you have covered in each lesson with what you already know.	Effort Regulation
	• Posting and replying to chats and forums created by yourself, your peers, and the course instructor enhances your knowledge absorption. Maintain this spirit as it will also help others learn.	Help-Seeking

4.3.2 System Conceptual Design and Architecture

This section describes two processes that involved translation of the needs analysis and conceptual scenarios into system conceptual design and system architecture.

First, the conceptual structure that represented information flow between the key components and data processing for the EDM plugin was visualized as illustrated in Figure 4.1.

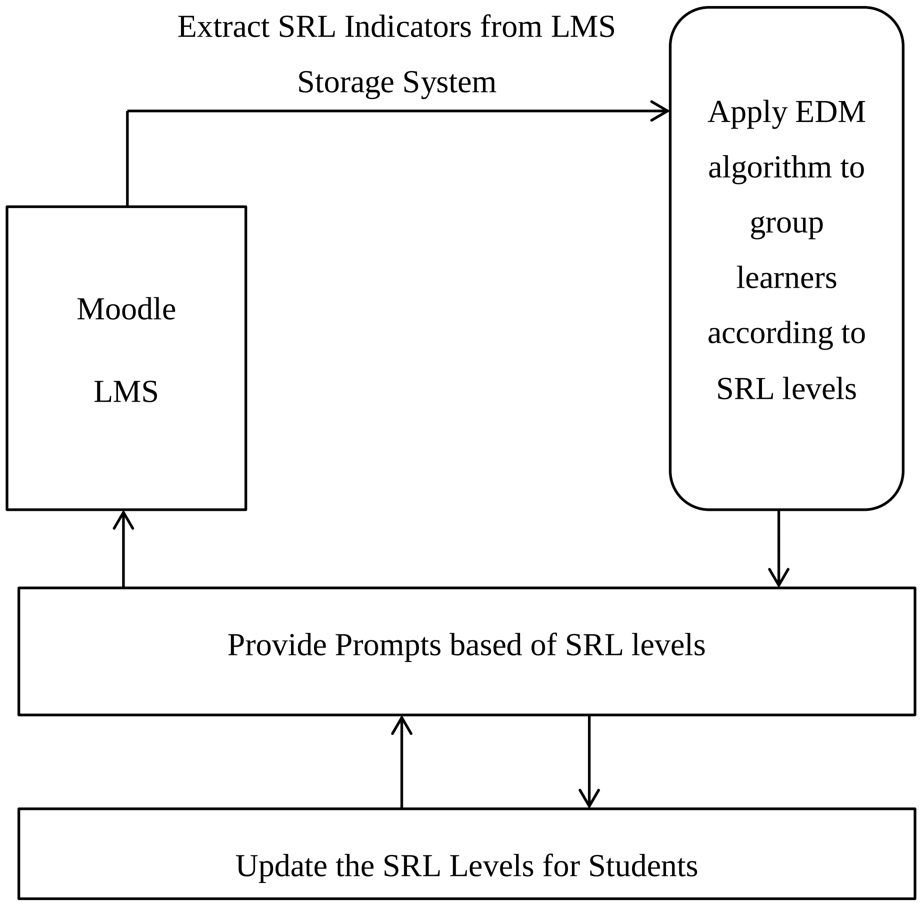


Figure 4.1: System Conceptual Design

The second task involved converting the conceptual design into an architectural design that represented information flow between the Moodle LMS, data storage, and EDM plugin. The clustering algorithm obtains raw data from Moodle LMS and forms clusters that are used to determine SRL levels for learners obtains and preprocesses Moodle LMS data before clustering the students according to their online learning behaviors as captured by the logs data. The clusters are defined by the EDM model. The classification algorithm integrated with the model uses the clustering information to place learners in relevant clusters. The clusters are then used to provide targeted prompts to the individual students. Based on the characteristics of the clusters of students, SRL interventions are provided in form of metacognitive prompts.

Figure 4.2 illustrates the architecture of the EDM interventions plugin.

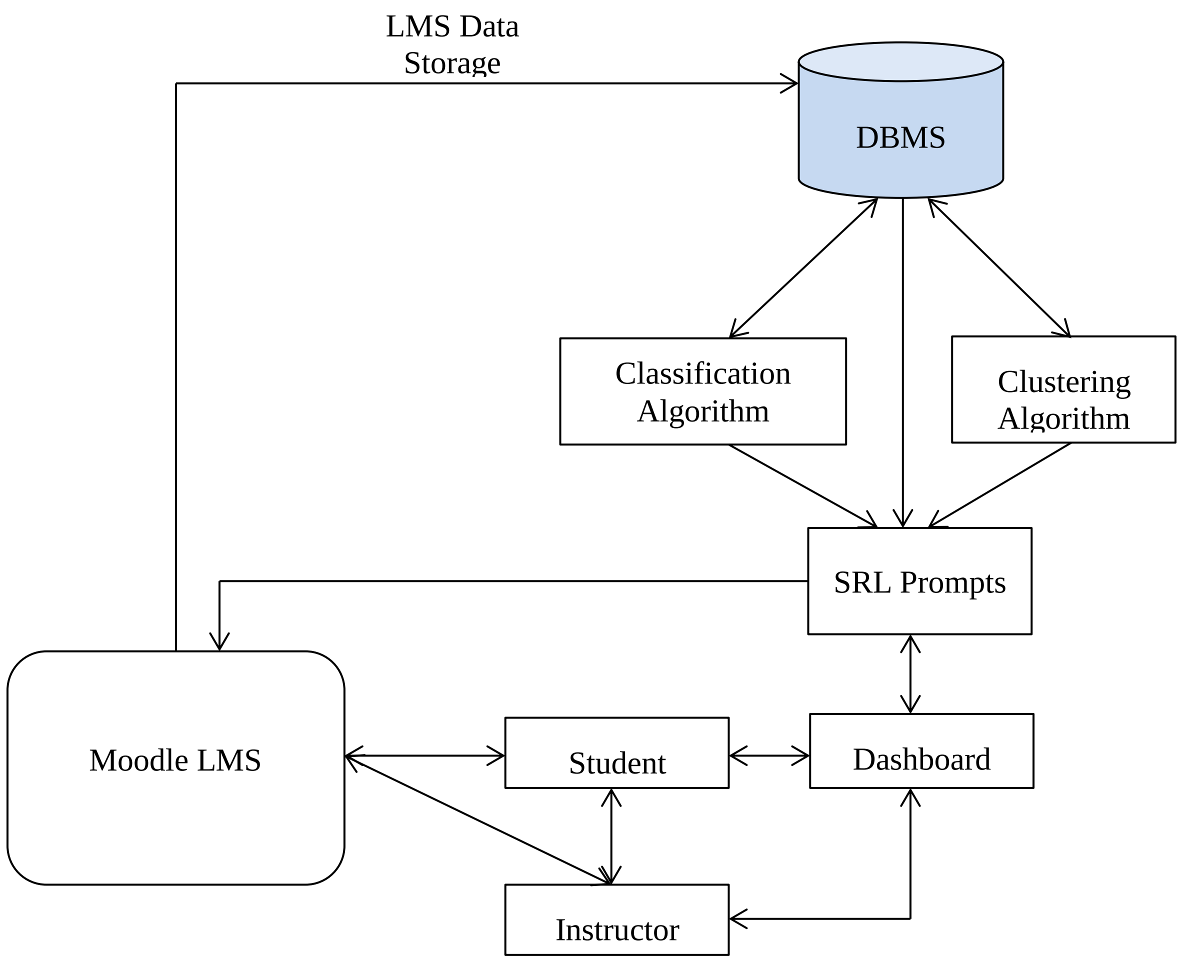


Figure 4.2: The Architecture of the EDM Model

4.3.3 System Development Process

The system development process involved the following processes;

a) Data collection and preprocessing
b) Development of clustering algorithm
c) Development of classification algorithm
d) Database design
e) Interface Design
f) Development of SRL prompting module
g) Integration of EDM model with Moodle LMS
h) Testing of the EDM Model

Figure 4.3 shows the flow activities carried out during the development of the EDM model.

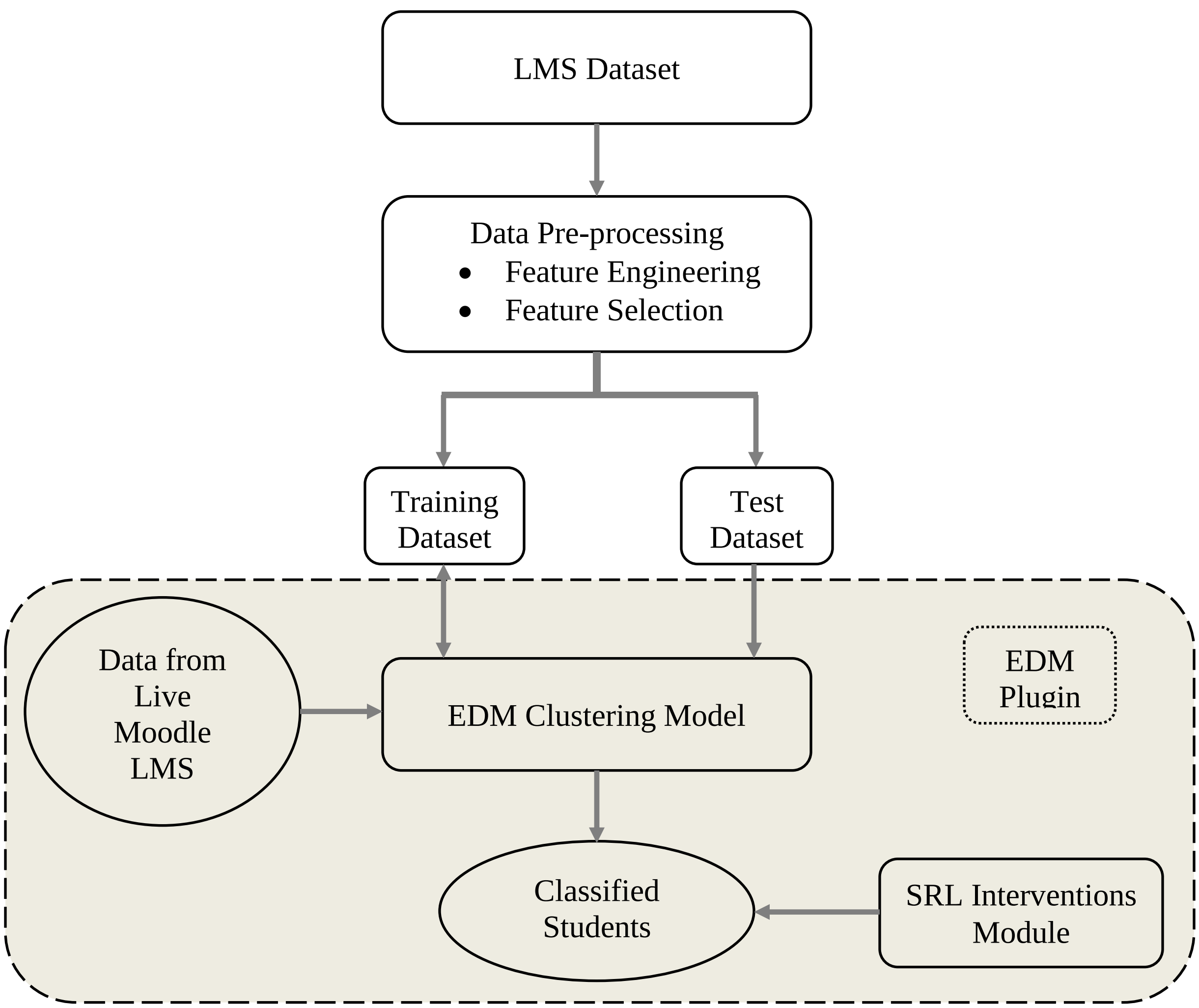


Figure 4.3: Educational Data Mining Model Development Process

a. Data collection and preprocessing

This process involved obtaining Moodle LMS log data for a previous course and identifying the SRL indicators that were under study. The extraction of features applied in modelling was achieved by manipulating students' log data using custom-made functions. The important columns in the log data used in the extraction of these features were the *student_id*, *component_column* and the corresponding *action*. To identify LMS features associated with a user, a for-loop function that was implemented in python was used to filter all log entries associated with that particular user. The next step was to obtain the number of log entries where that particular student engaged in a learning activity. For instance, to identify the number of assignments viewed, the function first checked for the assignment component (*mod_assign*") and evaluated the number of entries associated with viewing for this assignment. For this case, the *component* would be "*mod_assign*", while the action entry value would be "*view*".

Figure 4.4 shows the implementation of the "for_loop" function used to extract LMS features

```python
#Creating a DataFrame that is better structured for clustering
student_ids = df['userid'].unique()
def student_info(df):
    students_data = list()
    for num, i in enumerate(student_ids):
        student = df[df['userid'] == i]
        student = student[['userid', 'date', 'Component', 'Action']]
        student['Assignments_viewed'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='viewed')])
        student['Assignments_submitted'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='submitted')])
        student['Assignments_created'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='created')])
        student['Assignments_started'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='started')])
        student['Assignments_reviewed'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='reviewed')])
        student['Assignments_updated'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='updated')])
        student['Assignments_added'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='added')])
        student['Assignments_graded'] = len(student[(student['Component'] == 'mod_assign') & (student['Action']=='graded')])
        student['Quiz_viewed'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='viewed')])
        student['Quiz_submitted'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='submitted')])
        student['Quiz_created'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='created')])
        student['Quiz_started'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='started')])
        student['Quiz_reviewed'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='reviewed')])
        student['Quiz_updated'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='updated')])
        student['Quiz_added'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='added')])
        student['Quiz_graded'] = len(student[(student['Component'] == 'mod_quiz') & (student['Action']=='graded')])
        student['Forums_viewed'] = len(student[(student['Component'] == 'mod_forum') & (student['Action'] == 'viewed')])
        student['Forums_started'] = len(student[(student['Component'] == 'mod_forum') & (student['Action'] == 'started')])
        student['Forums_ended'] = len(student[(student['Component'] == 'mod_forum') & (student['Action'] == 'ended')])
        student['Chats_viewed'] = len(student[(student['Component'] == 'mod_chat') & (student['Action'] == 'viewed')])
        student['Charts_started'] = len(student[(student['Component'] == 'mod_chat') & (student['Action'] == 'started')])
        student['Resources_viewed'] = len(student[(student['Component'] == 'mod_resource') & (student['Action'] == 'viewed')])
        student['Date'] = student['date'].dt.date
        student = student.drop(columns=['Component', 'Action'], axis=1)
        students_data.append(student)
```

Figure 4.4: Implementation of the "for_loop" function use to obtain LMS features

The process of extracting the LMS features resulted in 22 parameters (features) for each student. Identify the most important features; first a histogram used to visualize the density of distribution of the 22 extracted features. The visualization revealed that some attributes were entirely populated with zero entries hence were insignificant. These features were considered outliers that would result in skewed results in subsequent steps of the analysis.

Figure 4.5 shows the histogram displaying the features and their respective densities as extracted from the raw log files.

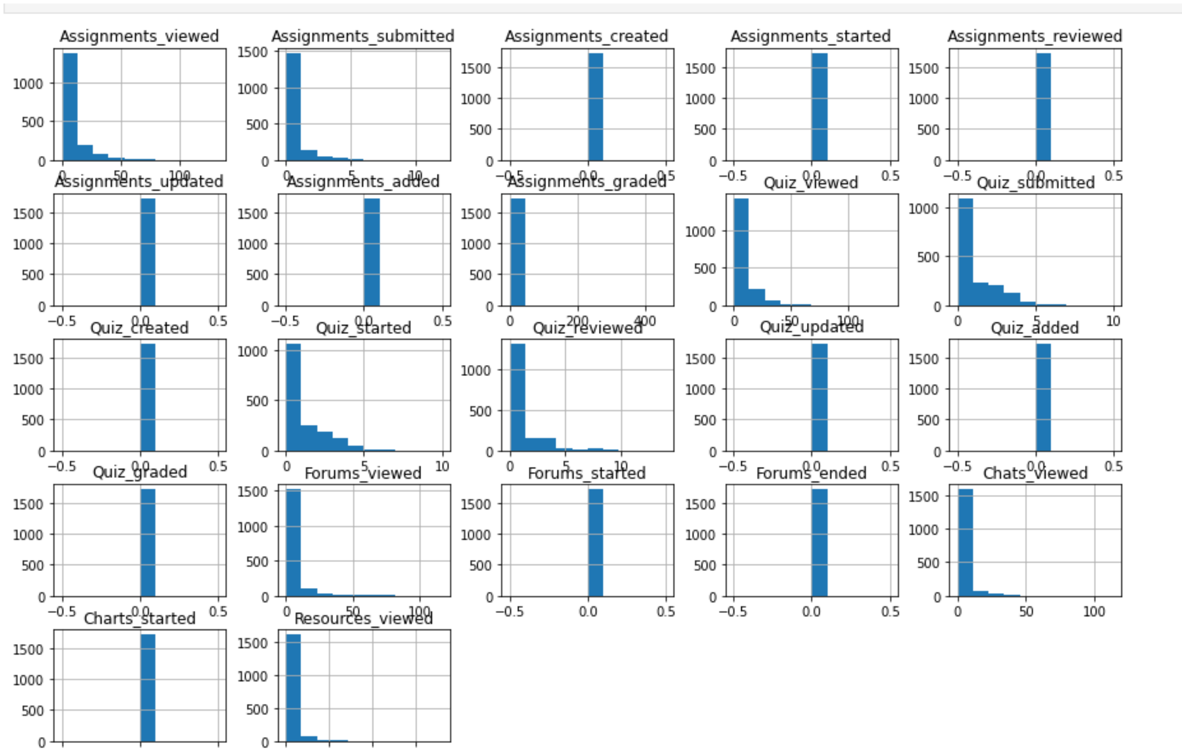


Figure 4.5: Histogram Visualizing the Density Distribution of the extracted LMS Features

To remain with important features, a custom filter was implemented which dropped all columns with an entry of zero. Nine parameters obtained from the preprocessing stage formed the input features to the subsequent machine learning clustering model that was developed.

The summary of the LMS features that were extracted and explored in this study is presented in Table 4.14.

Table 4.14: LMS Indicators for Measuring SRL Strategies

LMS Feature	Description
Assignments_submitted	The number of assignments submitted by the student
Assignments_viewed	The number of times the students clicked (to view) the assignment(s)
Quiz_started	The number of times the student interacted with the quiz without completing it
Quiz_submitted	The number of quizzes the student submitted
Quiz_viewed	The number of times the students clicked (to view) the quiz(s)
Quiz_reviewed	The number of times the student went back to the quiz(s) after completing it previously
Forum_viewed	The number of times the students clicked (to view) the discussion forum(s)
Page_views	The number of times the students clicked (to view) the content page(s)
Resources_viewed	The number of times the students clicked (to view) the course resources that included URLs, files, and videos

Figure 4.6 represents the correlation matrix for the training dataset. The figure shows how the features correlate and what feature(s) played a significant role in the classification of students. It can be observed that strongly dependent features that contributed to the clustering outcome were quiz_started, quiz_reviewed, quiz_submitted, assignments_viewed, and assignments_submitted.

	Assignments_viewed	Assignments_submitted	Quiz_viewed	Quiz_submitted	Quiz_started	Quiz_reviewed	Forums_viewed	Chats_viewed	Resources_viewed	userid
Assignments_viewed	1.000000	0.724243	0.486013	0.481639	0.488565	0.449104	0.482592	0.348031	0.263159	0.127107
Assignments_submitted	0.724243	1.000000	0.285885	0.363412	0.362998	0.313366	0.227937	0.185952	0.108997	0.082185
Quiz_viewed	0.486013	0.285885	1.000000	0.818313	0.832468	0.770318	0.555738	0.486342	0.209024	0.029726
Quiz_submitted	0.481639	0.363412	0.818313	1.000000	0.982868	0.876348	0.455732	0.361433	0.173863	-0.009037
Quiz_started	0.488565	0.362998	0.832468	0.982868	1.000000	0.867285	0.449284	0.360895	0.176543	-0.007557
Quiz_reviewed	0.449104	0.313366	0.770318	0.876348	0.867285	1.000000	0.460955	0.379823	0.179889	0.005787
Forums_viewed	0.482592	0.227937	0.555738	0.455732	0.449284	0.460955	1.000000	0.499026	0.285112	0.055130
Chats_viewed	0.348031	0.185952	0.486342	0.361433	0.360895	0.379823	0.499026	1.000000	0.200327	-0.000371
Resources_viewed	0.263159	0.108997	0.209024	0.173863	0.176543	0.179889	0.285112	0.200327	1.000000	0.008888
userid	0.127107	0.082185	0.029726	-0.009037	-0.007557	0.005787	0.055130	-0.000371	0.008888	1.000000

Figure 4.6: Correlation Matrix for LMS features used in Model developement

b. Development of clustering algorithm

The selection of the algorithm used in the clustering process was achieved after an empirical evaluation and comparison of multiple clustering techniques (Araka et al., 2022). In their study, Araka et al. (2022) empirically compares three commonly used clustering algorithms; K-Means, Agglomerative Hierarchical clustering, and Expectation-Maximization. The performance of algorithms was evaluated using the internal validation measures to identify the optimal algorithms and number of clusters using the clValid (Brock et al., 2008) and the NbClust (Charrad et al., 2014) R libraries.

Table 4.15 presents the optimal algorithm and cluster evaluation results.

Table 4.15: Experimental Evaluation of the Clustering Algorithms and Optimal Clusters

Algorithm	Validation measure	Number of Clusters							
		3	4	5	6	7	8	9	10
Agglomerative Hierarchical	Connectivity	**8.4552**	12.0135	20.7044	22.9044	25.9333	30.7552	43.1417	47.5131
	Dunn	0.0576	**0.0609**	0.0299	0.0299	0.0312	0.0356	0.0223	0.0250
	Silhouette	**0.7111**	0.7095	0.6116	0.6110	0.6024	0.5472	0.5054	0.5110
K-Means	Connectivity	12.9540	27.7000	42.8472	45.3774	47.5774	65.5869	74.4853	65.5829
	Dunn	0.0135	0.0075	0.0057	0.0131	0.0131	0.0082	0.0061	0.0185
	Silhouette	0.6571	0.5650	0.5443	0.5326	0.5316	0.4892	0.4615	0.4633
Expectation-Maximization	Connectivity	37.7675	47.3556	55.5512	62.8067	73.1829	86.7683	114.7929	128.2321
	Dunn	0.0009	0.0018	0.0017	0.0030	0.0023	0.0047	0.0026	0.0026
	Silhouette	0.5278	0.4491	0.4616	0.4709	0.4613	0.4062	0.3597	0.3551

Note. The optimal score value for Connectivity, which identifies the optimal number of clusters with the lowest score, and Dunn index and Silhouette which identifies the optimal number of clusters with the highest score are in bold (Brock et al., 2008).

The findings from the evaluation demonstrated that Agglomerative Hierarchical clustering is the best-performing algorithm compared to Expectation-Maximization and K-Means. The Agglomerative Hierarchical clustering algorithm is the best performing algorithm with the optimal score of 8.4552 for connectivity and 0.7111 for Silhouette measures that proposes 3 optimal clusters. However, the Dunn index proposes 4 optimal clusters. Moreover, to identify the optimal number of clusters the NbClust library was applied. The NbClust library provides 30 internal validation indices that allow simultaneous evaluation of algorithms to establish the optimal number of clusters for a given dataset (Charrad et al., 2014). From these 30 indices, seven proposed 3 as the optimal number of clusters, fifteen proposed 4 clusters, and two proposed 5 clusters. The rest of the indices, such as the Dindex and Hubert, gave graphical results. They also indicated that 4 clusters would be optimal. Based on the majority rule, it was concluded that the best number of clusters in the dataset would be 4.

The development of the clustering algorithm followed the conventional steps in the machine learning process that included data cleaning and pre-processing, feature engineering, modeling, and model evaluation. The primary data cleaning steps that were done were dropping columns with a maximum value of zero as well as label encoding the target column (target category). The final clustering model was developed to transform Moodle LMS data into four clusters using nine features.

Figure 4.7 shows the features that were used to classify students into the four clusters.

	Assignments_viewed	Assignments_submitted	Quiz_viewed	Quiz_submitted	Quiz_started	Quiz_reviewed	Forums_viewed	Chats_viewed	Resources_viewed
count	1640.000000	1640.000000	1640.000000	1640.000000	1640.000000	1640.000000	1640.000000	1640.000000	1640.000000
mean	5.632927	0.413415	5.556707	0.696951	0.723780	0.826220	3.579268	2.196341	1.900610
std	8.810701	0.908809	9.415717	1.166542	1.189714	1.642828	10.459594	6.812255	4.258488
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	8.000000	0.000000	8.000000	1.000000	1.000000	1.000000	2.000000	1.000000	2.000000
max	68.000000	9.000000	136.000000	7.000000	7.000000	14.000000	117.000000	114.000000	90.000000

Figure 4.7: Description of the Clustering Data from Moodle LMS

c. Development of Classification Algorithm

The classification algorithm was largely dependent on the previous SRL identification process and clustering process. The classification algorithm is used to place students into the right clusters progressively as every time updated log store data is generated. Through the identification of the SRL indicators, it was possible to transform the log data from Moodle LMS into a well-structured data file with clearly marked features. The implementation of the Agglomerative Hierarchical clustering algorithm in the clustering phase resulted in the development of target features (student cluster) for the data. Following the preparation of data, four classifiers were designed and compared to identify the champion model. The classifiers include logistic regression, decision tree classifier, random forest classifier, and Extreme-Gradient Boosting (XGBoost) classifier. Each classifier was compared using a stratified k-fold accuracy measure. A k-fold size of 10 splits was used.

Table 4.16 shows the summary of results obtained from the model comparison.

Table 4.16: Classification Model Comparison

Model Name	Accuracy	Standard deviation
Logistic regression	0.9871	0.008193
Decision tree classifier	0.9596	0.017642
Random forest classifier	0.9731	0.012869
XGBoost classifier	0.9766	0.012831

The comparison results indicate that the logistic regression model performs best. The accuracy score, as well as the standard deviation from the cross-validation process, are optimal compared to other models. However, although accuracy is a common metric for evaluating machine learning models, it is important to note that it can be equally misleading, especially in cases where the classes are imbalanced. To avoid this problem, a classification report for the logistic regression model was computed. The report, presented in Table 4.17 shows that the model has recall and precision scores that are significantly high, therefore ruling out any possibilities of model overfitting.

Table 4.17: Classification Report for the Logistic Regression Model

	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	316
1	1.00	0.99	0.99	81
2	1.00	1.00	1.00	13
Accuracy			1.00	410
Macro Average	1.00	1.00	1.00	410
Weighted Average	1.00	1.00	1.00	410

The design and delivery of intervention messages were designed and developed for each cluster. Clustering is done in real-time after analysis of students' log data. The clusters are continuously reviewed and used by the classifier to place the students in the new cluster to receive the appropriate SRL interventions. The EDM plugin is integrated into Moodle LMS backend where the clusters are defined. The integration was through the Moodle LMS local directory for plugins that provide an API to the main Moodle system. The frontend which is defined by PHP and linked to the backend using Python is connected with the classification algorithm that places students into appropriate clusters and provides the SRL interventions. The EDM model was designed to first fetch and analyze the un-clustered data stored within the Moodle LMS in the *mdl_logstore* table.

The classification of students into respective clusters, therefore, depends on the frequency and access to the parameters that were defined in the clustering model and implemented by the EDM model. After clustering, the generated data is used to train a classifier to place a student into the right cluster (*new_instance*). The parameters included *Assignments_submitted*, *Assignments_viewed*, *Quiz_started*, *Quiz_submitted*, *Quiz_viewed*, *Quiz_reviewed*, *Forum_viewed*, *Page_views*, *Resources_viewed*. After the clustering process, a new variable (cluster) is created and fed into the classifier. The classifier will use the new data to place the learner into the appropriate cluster.

Figure 4.8 shows a sample of records on how students were classified into respective clusters by the EDM model.

	Assignments_viewed	Assignments_submitted	Quiz_viewed	Quiz_submitted	Quiz_started	Quiz_reviewed	Forums_viewed	Chats_viewed	Resources_viewed	userid	Cluster_category
0	0	0	0	0	0	0	0	0	1	72310	0
1	1	0	0	0	0	0	0	0	0	85304	0
2	1	0	4	1	1	0	0	0	0	70964	0
3	8	0	8	2	2	2	0	1	2	3035	0
4	7	0	1	0	0	0	9	0	1	86001	0

Figure 4.8: Classification of Students into Clusters by the EDM model

The clustering takes place within the *model_lib.php* where the *main.py* file gives an output of the EDM model from the *new Classifier()* as an array. The data is then mapped to the *local_edm_student_cluster* table to each of the respective field parameters. The clustering is triggered as students are interacting within the LMS but the clustered data are displayed programmatically in a weekly time series. The clustered time is represented in epoch time.

d. Database design

This process involved the design, development, and integration of the EDM plugin database into the core Moodle database. The tables were defined as a schema in XML within the plugin and later integrated programmatically with Moodle LMS. Three schemas; *local_edm_student_cluster, local_edm_clusters, and local_edm_messages* were defined in the *install.xml file* of the plugin. The *local_edm_student_cluster* stores the clustered data from the ML model and the parameters defined by the clustering model. The *local_edm_clusters* stores the student's unique id and the clusters of the respective students. Lastly, the *local_edm_messages* contain information about the intervention messages.

```
<TABLE NAME="local_edm_messages" COMMENT="each record contains user created messages">
  <FIELDS>
    <FIELD NAME="id" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="true"/>
    <FIELD NAME="cluster_id" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="user_id" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="messagetext" TYPE="text" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="messagetype" TYPE="text" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="seen_status" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="time_created" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
  </FIELDS>
  <KEYS>
    <KEY NAME="primary" TYPE="primary" FIELDS="id"/>
  </KEYS>
</TABLE>
```

```
<TABLE NAME="local_edm_student_cluster" COMMENT="each record contains user cluster data">
  <FIELDS>
    <FIELD NAME="id" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="true"/>
    <FIELD NAME="assignments_viewed" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="assignments_submitted" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="quiz_started" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="quiz_submitted" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="quiz_reviewed" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="quiz_viewed" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="forums_viewed" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="chats_viewed" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="resources_viewed" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="time_clustered" TYPE="int" LENGTH="10" NOTNULL="true" DEFAULT="0" SEQUENCE="false"/>
    <FIELD NAME="cluster" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="false"/>
    <FIELD NAME="userid" TYPE="int" LENGTH="11" NOTNULL="true" DEFAULT="0" SEQUENCE="false" COMMENT="User id from table
  </FIELDS>
  <KEYS>
    <KEY NAME="primary" TYPE="primary" FIELDS="id"/>
  </KEYS>
</TABLE>
```

```
<TABLE NAME="local_edm_clusters" COMMENT="Stores all classified clusters">
  <FIELDS>
    <FIELD NAME="id" TYPE="int" LENGTH="10" NOTNULL="true" SEQUENCE="true"/>
    <FIELD NAME="user_id" TYPE="int" LENGTH="10" NOTNULL="false" SEQUENCE="false"/>
    <FIELD NAME="cluster" TYPE="int" LENGTH="10" NOTNULL="false" SEQUENCE="false"/>
  </FIELDS>
  <KEYS>
    <KEY NAME="primary" TYPE="primary" FIELDS="id"/>
  </KEYS>
</TABLE>
```

Foreign keys such as *user_id, userid, and cluster_id* are used to allow the identification of students and their various clusters. Also, the respective domain of every field within the tables is clearly defined and data types specified.

e. Interface Design

This process involved the design and development of the plugin dashboard. The dashboard has two interfaces. The first is the students' interface which presents the summary of the activities student engaged in previously, the intervention messages received, and the notifications sent via email. The code implementation of the interface defined the student's analytic panel where each respective student can view detailed analysis based on his interaction with the LMS. The students can view the recent messages sent to them via popups and emails, and the number of activities engaged on the LMS.

Figures 4.9 and 4.10 represent dashboard screenshots showing the summary of activities for each student and recent intervention messages sent to a student respectively.

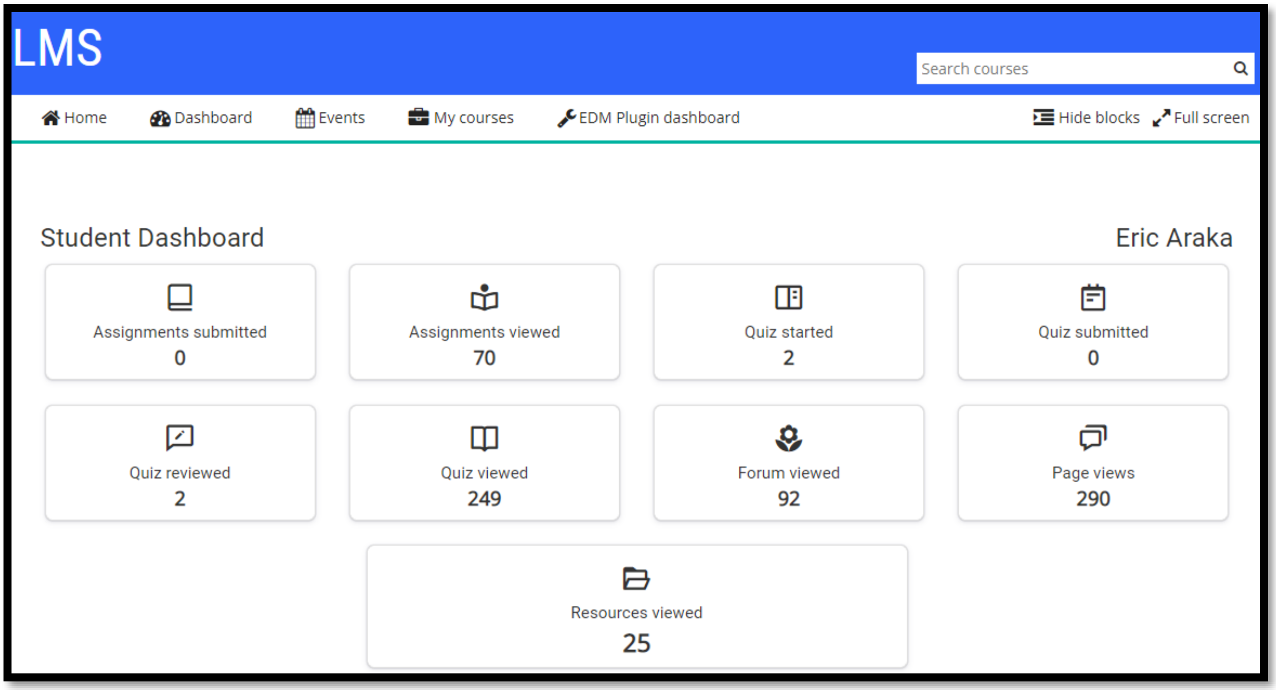


Figure 4.9: Student Dashboard showing the Summary of activities for each student

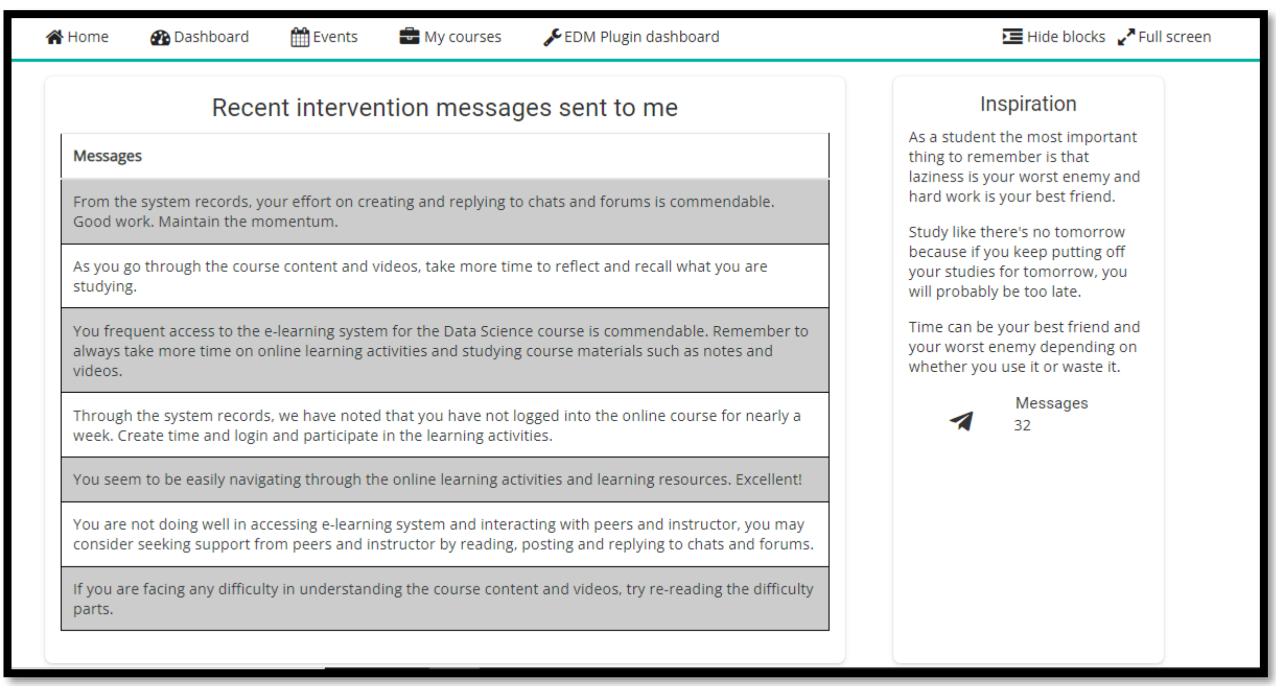


Figure 4.10: Student Dashboard showing Intervention Messages sent to a Student

Secondly, the instructor interface allows instructors access to clustered information on all the students. The instructor's interface defined the overview and general analysis of every student within the Moodle LMS. The dashboard displayed how students were distributed across each cluster. Other information such as the number of emails sent and the recent intervention messages are displayed on the instructor's interface. Charts based on the analytic data are also displayed to give a visual overview of the clustered data and the students in general. The distribution of all students among cluster activities is also displayed in a table format so that the instructor can be able to visualize the clustering of each student based on the cluster parameters is also provided. Figures 4.11, 4.12, and 4.13 show the screenshots obtained from the instructor dashboard.

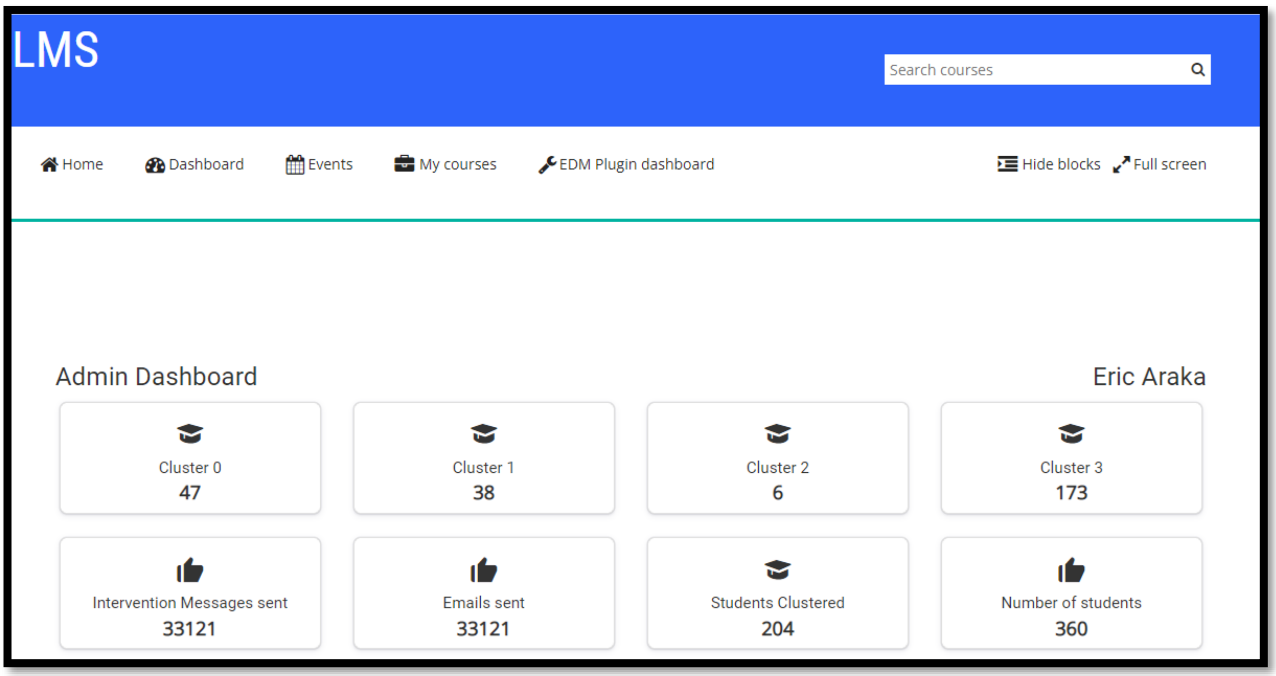


Figure 4.11: Instructor Dashboard showing Ditsribution of students among Clusters

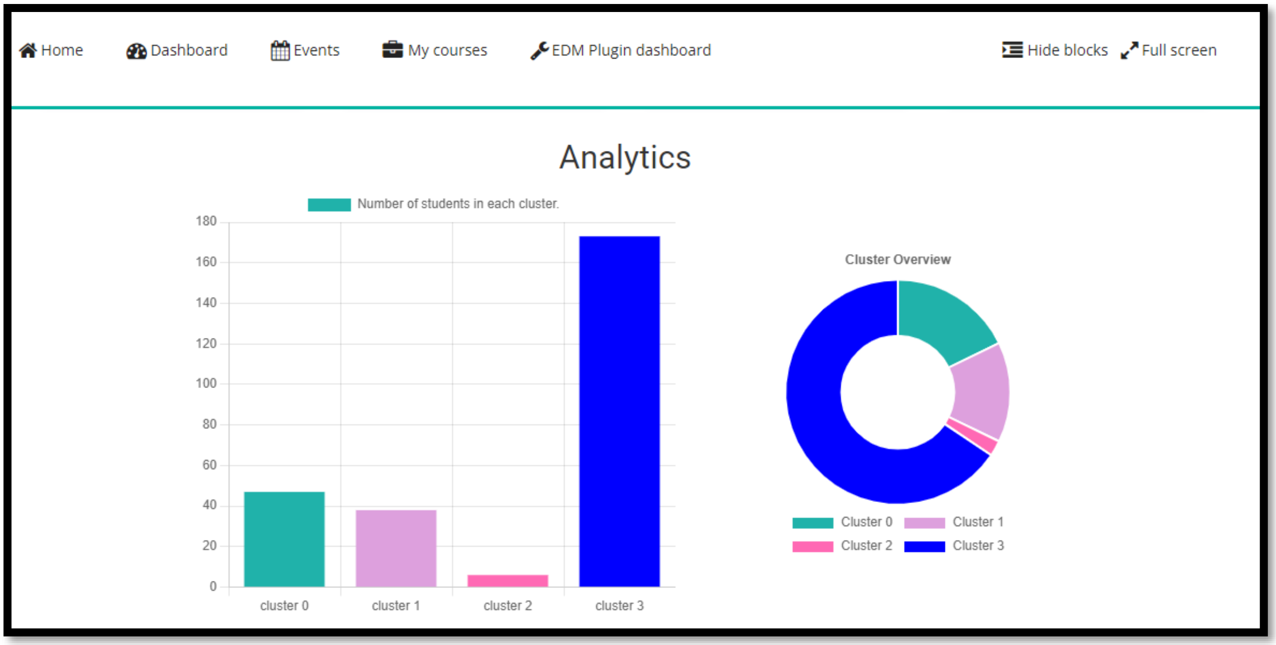


Figure 4.12: Dashboard Vizualization of number of Students in each Cluster

Home Dashboard Events My courses EDM Plugin dashboard Hide blocks Full screen

Distribution of students among clusters

StudentID	Assign_viewed	Assign_submitted	Q_started	Q_submitted	Q_viewed	Q_reviewed	Forums_viewed	Page_views	R_viewed	Cluster
0	0	0	0	0	0	0	183	0	0	3
2	0	0	0	0	0	0	12	0	3	0
6	0	0	0	0	2	0	0	0	0	0
12	70	0	2	0	249	2	92	13	25	3
13	0	0	0	0	3	0	51	0	58	1
14	421	0	0	0	0	0	18	0	5	2
16	18	1	4	4	104	3	37	5	19	3
17	39	1	2	2	60	1	50	10	6	3
18	134	4	2	2	118	1	289	37	60	3
19	94	1	2	2	43	1	28	0	22	3

Figure 4.13: Instructor Dashboard showing Distribution of Students among Clusters

f. Development of SRL Interventions Component

The intervention prompts were provided to learners within the LMS through pop-up windows and message containers within the EDM plugin. Moreover, to benefit the students who are always disengaged and rarely logins into LMS, the notification reminder messages were also sent to all the students enrolled in the course. The notification messages sent to students outside the LMS reinforced the intervention technique provided by the plugin to benefit those learners who are irregular visitors to the LMS. The prompts were spread out (staggered) in each lesson or learning episode for each student. The spreading out ensured that prompts helped students to improve their engagement behavior (Wong et al., 2020).

This process involved the development of the intervention messages module and how the email notification messages were implemented. The module is the implementation of the module sent out interventions message when students are interacting with the Moodle LMS in real-time. The implementation of the prompting module was achieved through the *local_edm_messages* table. The table stored the following information;

- The *user_id* references the student who has received a given message.
- The *cluster_id* references the cluster of a specific user based on the ML model of the plugin.
- The *messagetext* references the stored message intervention send to a specific student.
- The *messagetype* references the type of message level sent to a student as they vary based on the clustering.
- The *seen_status* references the status of the message to a specific student i.e. *whether a student has viewed the message or not*
- The *time_created* references the time when a specific message was sent to the student.
- The *id* refers to the unique identity of each record in the table.

After the development of the prompting module, a configuration of how each student received the intervention messages through the plugin within the Moodle LMS was defined by the *observer.php* file in the plugin codebase. The *observer.php* determines the specific instance where the student should be receiving the intervention message. The intervention messages were triggered when the students clicked on Moodle's *course_module* application programming interface. Specifically, the intervention messages are sent when a student accesses the course material within the LMS. When the interventions are triggered, the plugin confirms the cluster level of the student. On identifying the cluster level, a specific message based on the given cluster level is fetched and saved in the database at the logout episode. When the student accesses the course material, the saved message is then retrieved and displayed to the student via a popup message.

The popup message was developed in a creative way such that students were not able to block the popup via a browser or the LMS. This meant that the deliverability of the message was guaranteed at all times. Figure 4.14 shows a sample intervention message sent to a student within Moodle LMS. Additionally, the students' email reminders especially for those who could rarely access the LMS. The emails were scheduled to be sent twice weekly (Monday and Wednesday) at 9 am. The email script was configured using Gmail SMTP via PHPMailer library using composer. The script was set via the LMS server cronjob to enable effective scheduling and reminders to students.

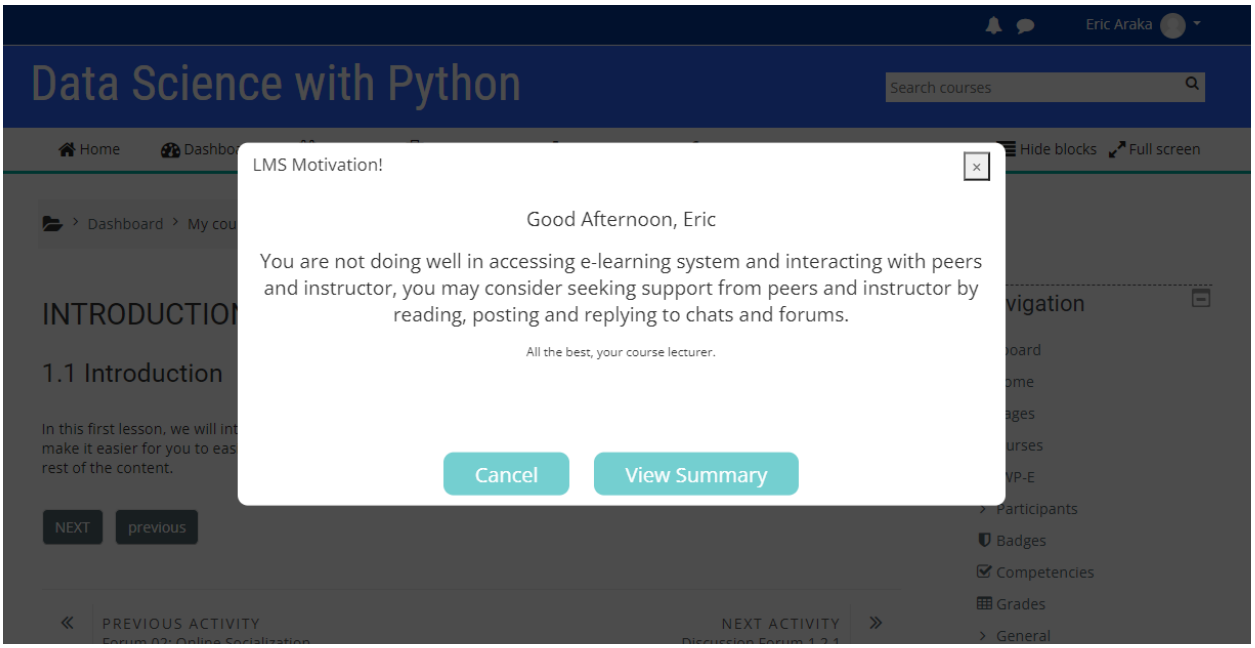


Figure 4.14: Sample Pop-up Message within Moodle LMS

g. Integration of EDM Model into Moodle LMS

This process describes how the EDM plugin was integrated into the Moodle LMS. Moodle LMS is structured in a modular manner that allows the extension of the LMS codebase and functionality. Moodle's local directory allows the addition of Moodle functionality in form of plugins. This is the functionality that was exploited in this study. All the source files of the EDM plugin are kept under the local directory of the LMS. The Moodle LMS automatically detects the plugin and then installs the plugin to the Moodle LMS codebase. During installation, Moodle LMS reads the *install.xml* schema to define and add the tables within its existing database schema. This process can be done via the CLI *command-line interface or the Moodle LMS interface.*

h. Evaluation of the EDM Model

The EDM model was evaluated based on three metrics; (a) classification of students into clusters, (b) display of the analytics dashboard that displayed a summary of learning activities for each student, and (c) the provision of interventions to the students. Table 4.18 shows the distribution of students among the clusters and the summary of learning activities for each student in the second week of learning.

Table 4.18: Summary of Students' Learning Activities and their Clusters during System Testing

StudentID	Assignment Viewed	Assignment Submitted	Quiz Started	Quiz Submitted	Quiz Viewed	Quiz Reviewed	Forums Viewed	Page Views	Resource views	Cluster Placed
455	16	0	0	0	0	0	0	0	1	0
456	0	0	0	0	2	0	0	0	0	0
458	0	0	0	0	0	0	0	0	0	0
461	0	0	0	0	0	0	0	0	0	0
462	0	0	0	0	0	0	11	0	4	0
464	0	0	0	0	2	0	0	0	16	0
475	0	0	0	0	0	0	0	0	1	0
476	0	0	0	0	0	0	0	0	5	0
477	0	0	0	0	0	0	4	1	1	0
487	0	0	0	0	0	0	0	0	0	0
499	0	0	0	0	0	0	0	0	3	0
503	0	0	0	0	0	0	1	0	0	0
516	0	0	0	0	0	0	1	0	5	0
519	0	0	0	0	0	0	0	0	0	0
528	0	0	0	0	0	0	1	0	0	0
529	0	0	0	0	0	0	2	0	0	0
530	0	0	0	0	0	0	0	0	1	0
534	0	0	0	0	0	0	1	0	2	0
537	0	0	0	0	0	0	0	0	2	0
538	0	0	0	0	0	0	0	0	0	0
541	0	0	0	0	0	0	0	0	0	0
542	0	0	0	0	0	0	0	0	0	0
545	0	0	0	0	0	0	0	0	0	0
547	0	0	0	0	0	0	2	0	0	0
553	0	0	0	0	0	0	2	0	2	0
556	0	0	0	0	0	0	3	0	3	0

557	6	0	0	0	0	0	0	0	0	0
558	0	0	0	0	0	0	0	0	2	0
560	0	0	0	0	0	0	0	0	2	0
563	0	0	0	0	0	0	6	1	4	0
569	0	0	0	0	0	0	2	0	1	0
571	0	0	0	0	0	0	0	0	1	0
576	0	0	0	0	0	0	0	0	1	0
579	0	0	0	0	0	0	0	0	2	0
583	0	0	0	0	0	0	2	0	0	0
585	0	0	0	0	0	0	0	0	15	0
590	1	0	0	0	0	0	13	0	3	0
591	0	0	0	0	0	0	3	0	0	0
595	0	0	0	0	0	0	0	0	1	0
596	0	0	0	0	0	0	0	0	2	0
600	0	0	0	0	0	0	5	0	1	0
601	0	0	0	0	0	0	0	0	2	0
602	0	0	0	0	0	0	3	0	0	0
610	0	0	0	0	0	0	0	0	1	0
611	0	0	0	0	0	0	0	0	0	0
612	0	0	0	0	0	0	1	0	0	0
614	0	0	0	0	0	0	0	0	0	0
615	0	0	0	0	0	0	0	0	0	0
616	0	0	0	0	0	0	0	0	0	0
620	0	0	0	0	0	0	0	0	0	0
623	0	0	0	0	0	0	1	0	3	0
626	0	0	0	0	0	0	1	0	0	0
628	0	0	0	0	0	0	8	1	1	0
630	0	0	0	0	0	0	0	0	0	0

4.4 Experimental Findings

This section presents the findings from the experiment and post studies that were carried out. First, the pretest and posttest scores from both the experimental and control groups are presented. Second, the Moodle LMS log files obtained from the experimental and control systems and EDM plugin log data obtained from the experimental system were analyzed and the results are described. Last, the results from the post-study structured survey and semi-structured interview carried out on the experimental group are presented.

4.4.1 Pretest and Posttest Results

The MSLQ scale was used to measure students' levels of SRL before and after the online learning course. The MSLQ questionnaire contained 32 items that included control of learning beliefs (4), self-efficacy for learning and performance (8), rehearsal (4), time and study environment (8), effort regulation (4), and help-seeking (4). The MSLQ is a likert scale tool and therefore Cronbach's alpha values were first computed to establish the reliability of pretest and post-test scores. The pretest was carried out to test for homogeneity of the two groups of students who participated in the study. Post-test was carried out to test for any significant difference between the experimental and control groups.

a) Pretest Results

Before the start of the online course, students completed the pretest questionnaire to measure the SRL score before the start of the experiments. Cronbach's alpha value for scores; was 0.65 for the experimental group and 0.66 for the control group, indicating an acceptable reliability level was obtained. The Shapiro-Wilk test for normality indicated that the data were not normally distributed and therefore non–parametric test using the Mann-Whitney U test was performed to establish any significant differences in SRL skills for the two groups.

Table 4.19 presents the pretest results from the Mann-Whitney U test that was conducted.

Table 4.19: Pretest Results for SRL Skills

SRL Strategy	Experimental (N=387)		Control (N=387)			
	M	*SD*	*M*	*SD*	Statistic	p-value
Control of Learning Beliefs	5.92	0.71	5.88	0.71	76834.5	0.47
Self-Efficacy for Learning & Performance	4.30	0.85	4.33	0.79	74657.0	0.94
Rehearsal	2.90	0.30	2.88	0.33	76428.5	0.35
Time & Study Environment	4.24	0.77	4.22	0.83	74649.5	0.94
Effort Regulation	2.75	0.94	2.79	0.98	74343.0	0.85
Help-Seeking	3.73	0.62	3.73	0.58	75277.5	0.86

From the Mann-Whitney U test statistic score and p-values for the SRL strategies, there was no significant difference in the scores between the experimental and control groups indicating that there was uniformity across participants' characteristics.

b) Posttest Results

First, the Shapiro-Wilk test for the distribution of the scores was conducted and the result indicated that the data were not normally distributed therefore non–parametric test using the Mann-Whitney U test was performed to establish any significant difference in SRL strategies for the two groups. Second, the Cronbach's alpha value for the scores was 0.75 for the experimental group and 0.67 for the control group, indicating an acceptable reliability level.

Table 4.20 presents the posttest results from the Mann-Whitney U test that was conducted.

Table 4.20: Posttest Results for SRL Skills

SRL Strategy	Experimental (N=124)		Control (N=124)			
	M	*SD*	*M*	*SD*	Statistic	p-value
Control of Learning Beliefs	5.23	0.83	5.37	0.77	6873.0	0.16
Self-Efficacy for Learning & Performance	3.85	0.94	3.88	0.96	7619.0	0.90
Rehearsal	2.62	0.55	2.70	0.48	7203.5	0.29
Time & Study Environment	3.68	0.95	3.77	0.94	7263.0	0.43
Effort Regulation	2.76	0.98	2.44	0.91	9051.0	0.01
Help-Seeking	3.15	1.04	3.13	1.16	7537.5	0.77

The Mann-Whitney U test results indicate that effort regulation scores for the experimental group are significantly greater compared to the control group, *p-value=0.01* which is statistically significant as $p<0.05$. The Mann-Whitney U test statistic and p-values for control of learning beliefs, self-efficacy for learning and performance, rehearsal, time and study environment, and help-seeking reveal that there was no significant difference in SRL strategies between the experimental group and control group.

4.4.2 Results from Moodle LMS Data

Six hundred and ninety-six students registered and were enrolled for the online data science with python course. Seventy-two participants were excluded from the LMS log data since they did not access the LMS after enrollment up to the end of the course. Three hundred and ninety-three students did not complete the mandatory assessments at the end of the course, therefore, were also excluded from the final analysis that involved grade book analysis for learning outcomes.

A total of 231 students completed the online course as presented in Figure 4.15.

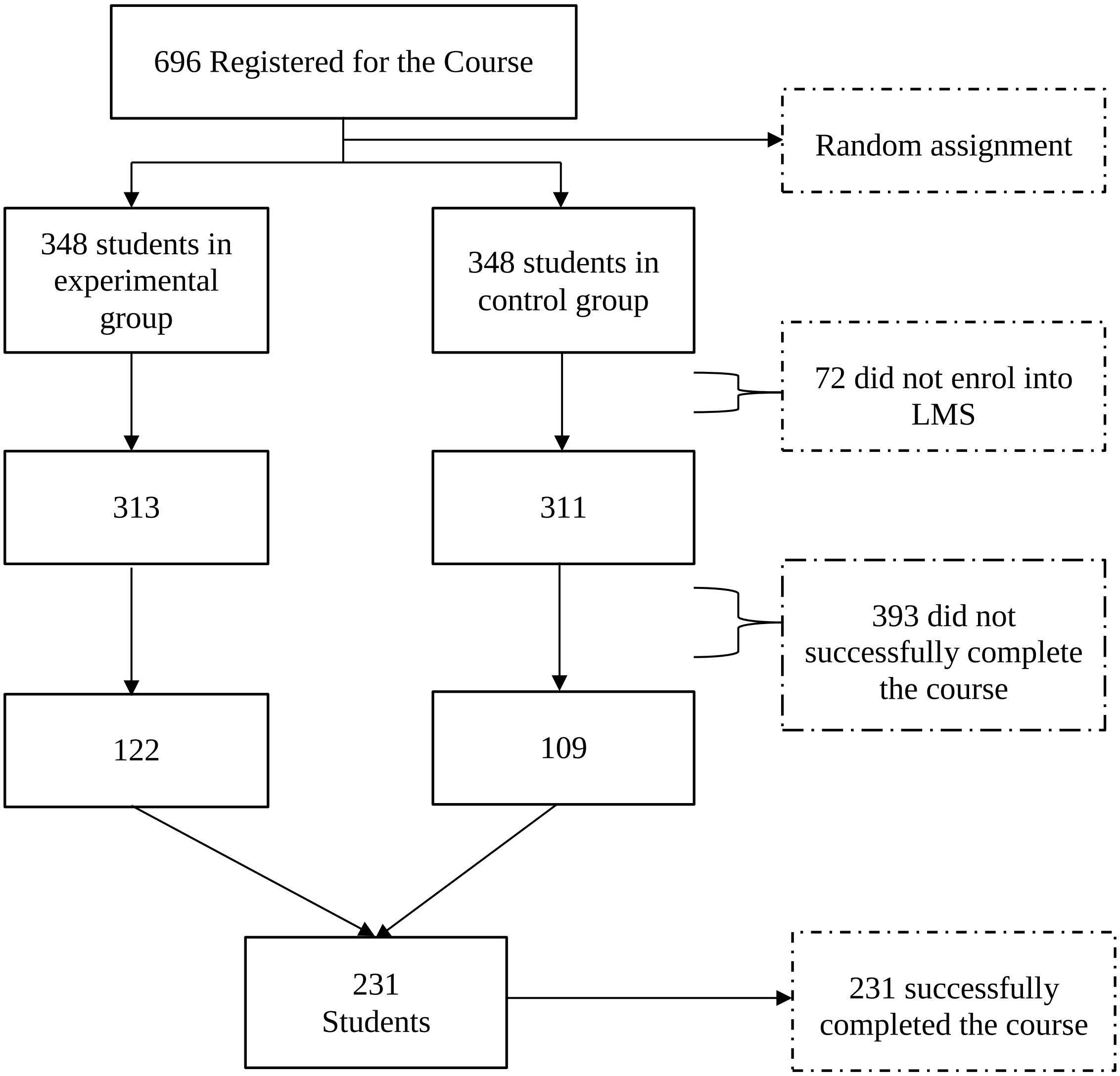


Figure 4.15: Students' Progression during the Experiment

The Moodle LMS log files from both control and experimental systems were obtained and uploaded into the python programming environment for preprocessing and feature selection. Preprocessing involved removing logs for guests, lecturers, and administrator logs and transforming the dataset to obtain the features that were initially used to train the clustering algorithm.

a. Results from Moodle LMS Data

The LMS data obtained from both experimental and control groups were visualized and analyzed for inferences on SRL strategies employed during the learning process. Figures 4.16 and 4.17 present the number of LMS activities for various events the students engaged in the experimental and the control groups respectively.

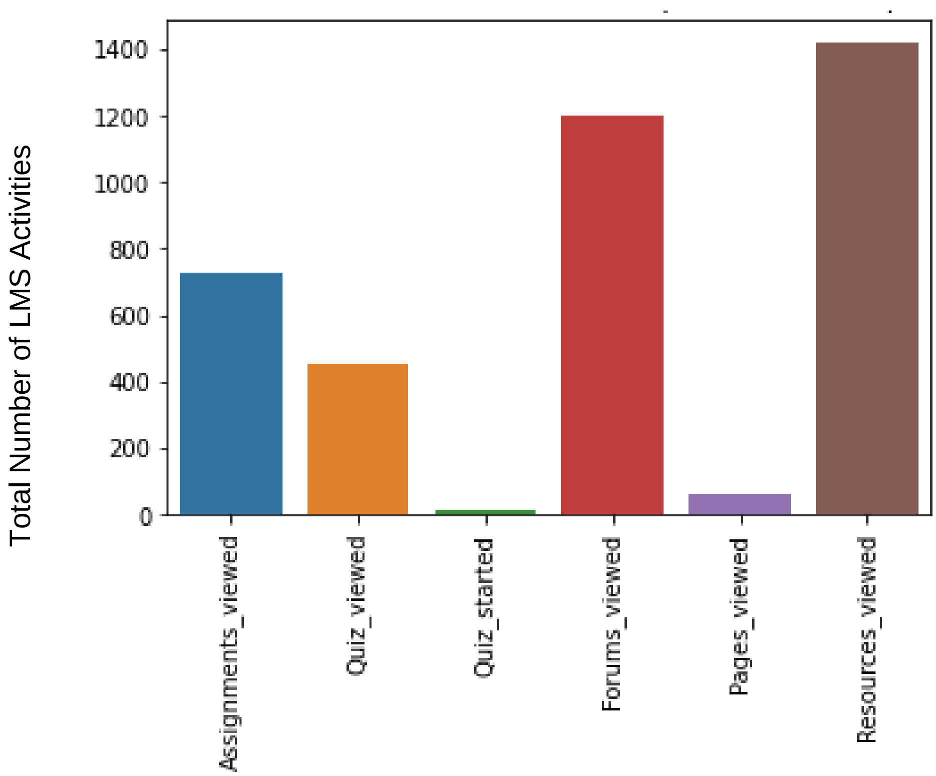


Figure 4.16: Distribution of LMS Activities for Students in the Experimental Group

The total number of LMS activities as identified from the LMS log data include assignments viewed = 723, quizzes viewed = 489, quizzes started = 12, forums viewed = 1202, page views = 66 and resources viewed = 1475 for experimental group compared to assignments viewed = 578, quizzes viewed = 427, quizzes started = 18, forums viewed = 816, page views = 75 and resources viewed = 615 for control group as presented in Figure 4.16 and Figure 4.17 respectively.

Two features that include assignments and quizzes submitted were excluded from the analysis since the number of submissions for quizzes and assignments was the same for all students in both experimental and control groups.

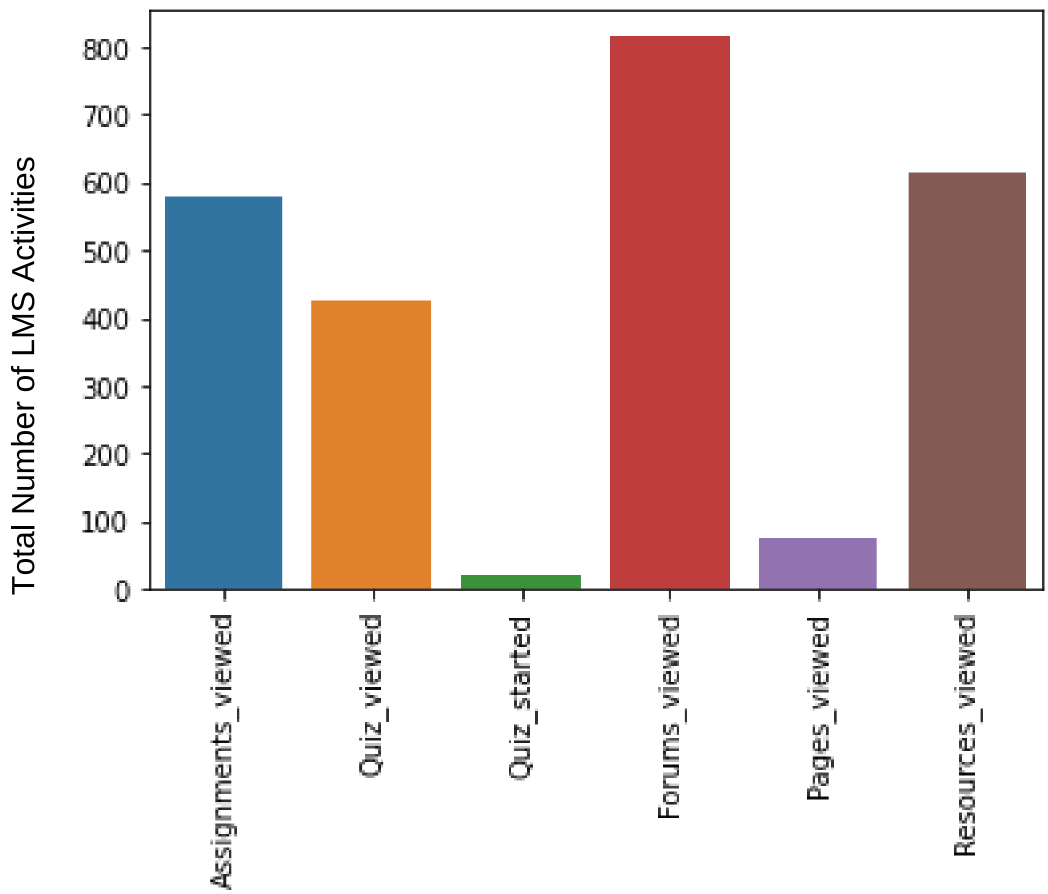


Figure 4.17: Distribution of LMS Activities for Students in the Control Group

To test for any significant difference in the LMS activities of the students between the experimental and the control groups, an independent t-test was carried out, and the results are presented in Table 4.21.

Table 4.21: T-test Results for LMS activities reflecting Learner Engagement

LMS Activity	Experimental (N=109)		Control (N=109)			
	M	*SD*	*M*	*SD*	Statistic	p-value
Assignments viewed	6.61	4.95	5.30	3.79	2.18	0.03
Quizzes viewed	4.37	4.38	3.92	3.85	1.02	0.31
Quiz started	0.10	0.35	0.17	0.37	-1.10	0.27
Forums viewed	10.77	10.96	7.49	7.27	2.75	0.01
Pages viewed	0.59	1.08	0.68	1.20	-0.52	0.60
Resources viewed	13.03	11.64	5.64	4.51	6.47	0.00

When the t-distribution table value of t (216) = 1.984 at an alpha value of 0.05 is compared to the computed t-statistic values for LMS activities (assignments viewed = 2.18; forums viewed = 2.75; and resources viewed = 6.47), the computed t-values are greater than the t-distribution table value. Additionally, the p-values for the LMS activities are less than the alpha level; $p<0.05$. We therefore conclude that there was a significant difference in assignments viewed, forums viewed and resources viewed between the experimental and control groups since $p<0.05$. The students in the experimental group had more views on assignments, forums, and resources compared to the control group.

b. SRL strategies Identified from Moodle LMS data

The LMS interaction data provided the actual events undertaken by students during the online course. Each event indicator was mapped to the corresponding SRL strategy. For each SRL strategy (unobserved features) a set of LMS indicators (observed features) were identified. Where more than one indicator corresponding to one strategy was identified, an average count was obtained. Seven parameters were identified from LMS and mapped to five SRL strategies based on Pintrich's (2000) model. Students' interaction data were used to detect online learning behavior by relating the indicators and the corresponding SRL strategy.

The summary of the SRL strategies identified from LMS data is presented in Table 4.22.

Table 4.22: SRL Strategies Identified from Moodle LMS data

LMS Feature	Description	Corresponding SRL Strategy
Assignments_viewed	The number of times the students clicked (to view) the assignment(s)	Rehearsal Time on Task
Quiz_started	The number of times the student interacted with the quiz without completing it	Effort Regulation
Quiz_viewed	The number of times the students clicked (to view) the quiz(s)	Rehearsal Time on Task
Quiz_reviewed	The number of times the student went back to the quiz(s) after completing it previously	Effort Regulation Rehearsal
Forum_viewed	The number of times the students clicked (to view) the discussion forum(s)	Help-Seeking
Page_views	The number of times the students clicked (to view) the content page(s)	Time of task Time Management
Resources_viewed	The number of times the students clicked (to view) the course resources that included URLs, files, and videos	Time of task Time Management

Table 4.23 presents the LMS indicators and the corresponding SRL Strategies. Two SRL strategies and five SRL constructs were identified from LMS log data.

Table 4.23: Transformation of LMS Activities into Corresponding SRL Strategies

SRL Strategy	SRL Construct	LMS Indicator (Average Count on combined variables)
Cognition	Time on Task	Assignments viewed, Quizzes viewed, Page views, and Resources viewed
	Rehearsal	Assignments viewed, Quizzes viewed, and QR=Quizzes reviewed
Behavior	Time Management	Pageviews and Resources viewed
	Effort Regulation	Quizzes Submitted, and Quizzes Reviewed
	Help-Seeking	Forums Viewed

The LMS data representing each SRL strategy for both experimental and control groups were visualized and analyzed.

Figures 4.18 and 4.19 present the distribution of SRL strategies in experimental and control groups respectively.

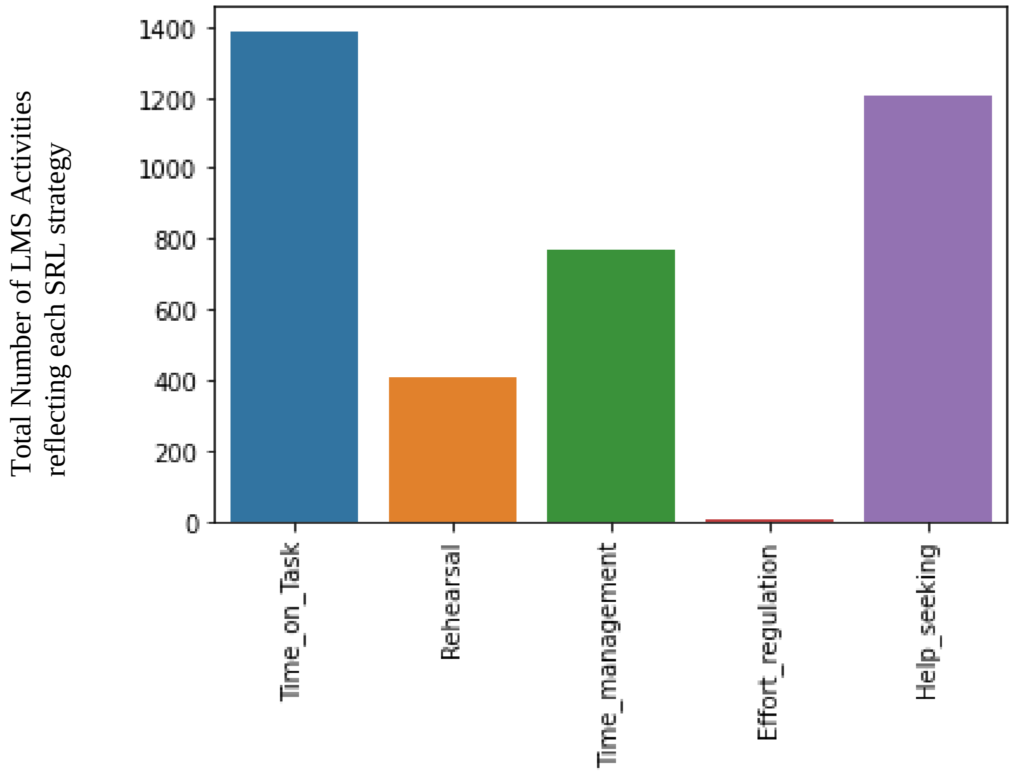


Figure 4.18: Distribution of SRL Strategies for the Experimental Group

The total number of LMS activities reflecting each SRL strategy as identified from the LMS log data include Time on Task = 1387.33, Rehearsal = 404.00, Time Management = 770.50, Effort Regulation = 6.00, and Help Seeking = 1202.00 for experimental group compared to Time on Task = 745.00, Rehearsal = 335.00, Time Management = 345.00, Effort Regulation = 9.00, Help Seeking = 816.00 for control group as presented in Figure 4.18 and Figure 4.19 respectively.

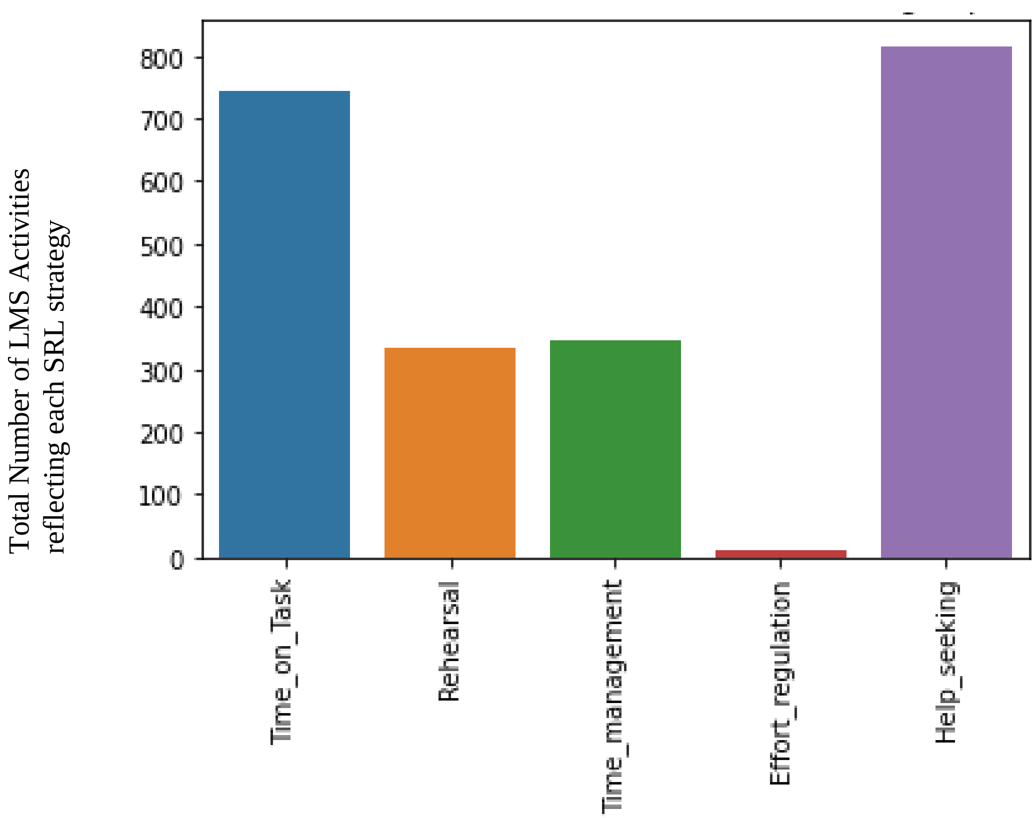


Figure 4.19: Distribution of SRL Strategies for the Control Group

To evaluate whether the interventions provided by the EDM plugin promote SRL strategies, an independent t-test between the experimental group (with interventions) and control group (without interventions) was carried out and the result is presented in Table 4.24.

Table 4.24: T-Test Results for SRL Strategies identified from LMS Data

SRL Strategy	SRL Construct	Experimental (N=109)		Control (N=109)			
		M	SD	M	SD	Statistic	p-value
Cognition	Time of Task	12.73	9.14	6.83	4.25	6.10	0.00
	Rehearsal	3.71	2.50	3.07	2.04	2.04	0.04
Behavior	Time Management	7.07	6.11	3.17	2.47	6.18	0.00
	Effort Regulation	0.06	0.18	0.08	0.19	-1.10	0.27
	Help-seeking	11.03	11.33	7.48	7.27	2.74	0.01

When the t-distribution value of t (216) = 1.984 at an alpha value of 0.05 is compared to the computed t-statistic values for SRL strategies (time on task = 6.10; rehearsal = 2.04; time management = 6.18; and help-seeking =2.74), the computed t-values are greater than the t-distribution table value. Additionally, the p-values for the SRL strategies are less than the alpha level; $p<0.05$. We therefore reject the null hypothesis and fail to reject the alternative hypothesis. The independent t-test results indicate that time on task, rehearsal, time management and help-seeking scores for the experimental group are significantly different to those of the control group with a p-value ranging between *0.00* to *0.04* which is statistically significant as $p<0.05$. The t-test statistic and p-values for effort regulation reveal that there was no significant difference in SRL strategy (effort regulation) between the experimental group and control group. Given this, it can be observed that the EDM-based interventions provided to the students in the experimental group via visualized feedback, prompts and email notifications were able to reinforce and improve students' SRL skills as compared to their counterparts in the control group who only received randomized instructor-led interventions on the announcements component of Moodle LMS as is conventionally done by online instructors.

c. Results from the EDM plugin

The EDM plugin captured the number of students in each cluster at a given instance and the summary of the LMS activities they had participated in.

Table 4.25 presents sample records from the EDM plugin showing the distribution of students among clusters.

Table 4.25: Sample Records on the Distribution of Students among Clusters

StudentID	Assignments viewed	Assignments submitted	Quiz started	Quiz submitted	Quiz viewed	Quiz reviewed	Forums viewed	Page views	Resources viewed	Cluster
12	70	0	2	0	252	2	92	13	25	3
13	0	0	0	0	3	0	51	0	58	1
14	470	0	1	0	32	1	18	0	6	2
16	18	1	4	4	104	3	37	5	19	3
17	52	2	3	3	125	2	50	10	6	3
18	134	4	3	2	170	1	289	37	61	3
19	134	3	3	3	139	2	30	0	30	3
20	38	2	2	2	71	1	35	12	12	3
21	9	0	1	1	38	1	11	0	3	3
22	51	2	2	2	45	1	8	1	37	3
23	12	0	2	2	50	1	9	0	4	3
24	21	2	2	2	40	1	3	0	3	3
25	0	0	0	0	0	0	10	0	2	0
27	63	2	4	4	201	45	50	1	44	3
29	0	0	0	0	0	0	1	0	2	0

For every student login instance, the LMS data for that student was queried and the students were classified in the cluster they belonged to as defined by the EDM model parameters. This enables a lecturer to be able to observe how a student progresses from one cluster to another.

An example of cluster evolution over time is illustrated in Figure 4.20 for a student with *user_id* 401 and how the student evolved from cluster 0 to cluster 3 during the online course

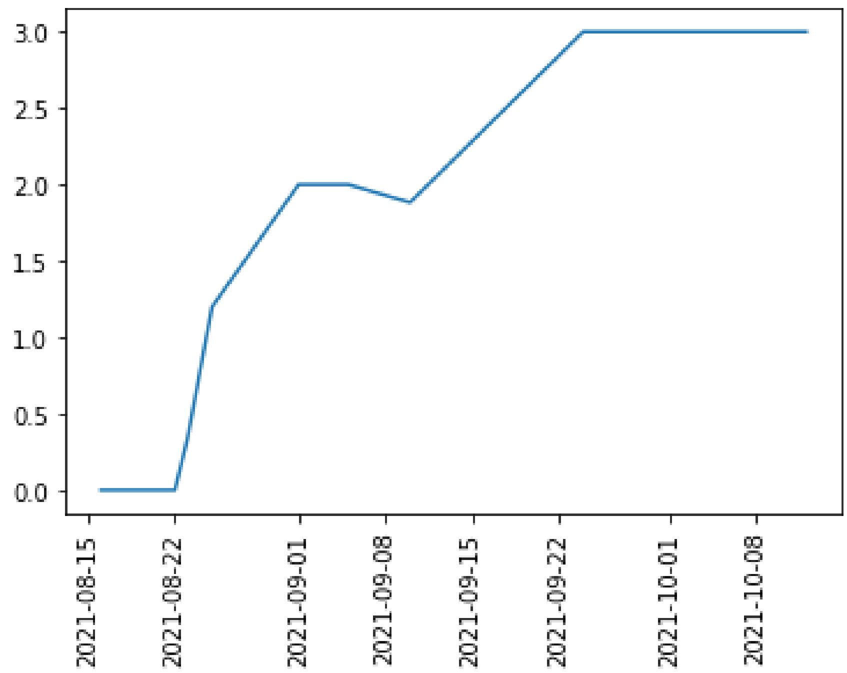


Figure 4.20: An Example of a student evolving from Cluster 0 to Cluster 3

Overall, Figure 4.21 shows how the number of students in each cluster changed over time during the whole course. As illustrated the number of students in cluster 3 kept on increasing while the number of students in clusters 0, 1, and 2 kept reducing to almost zero at the end of the course.

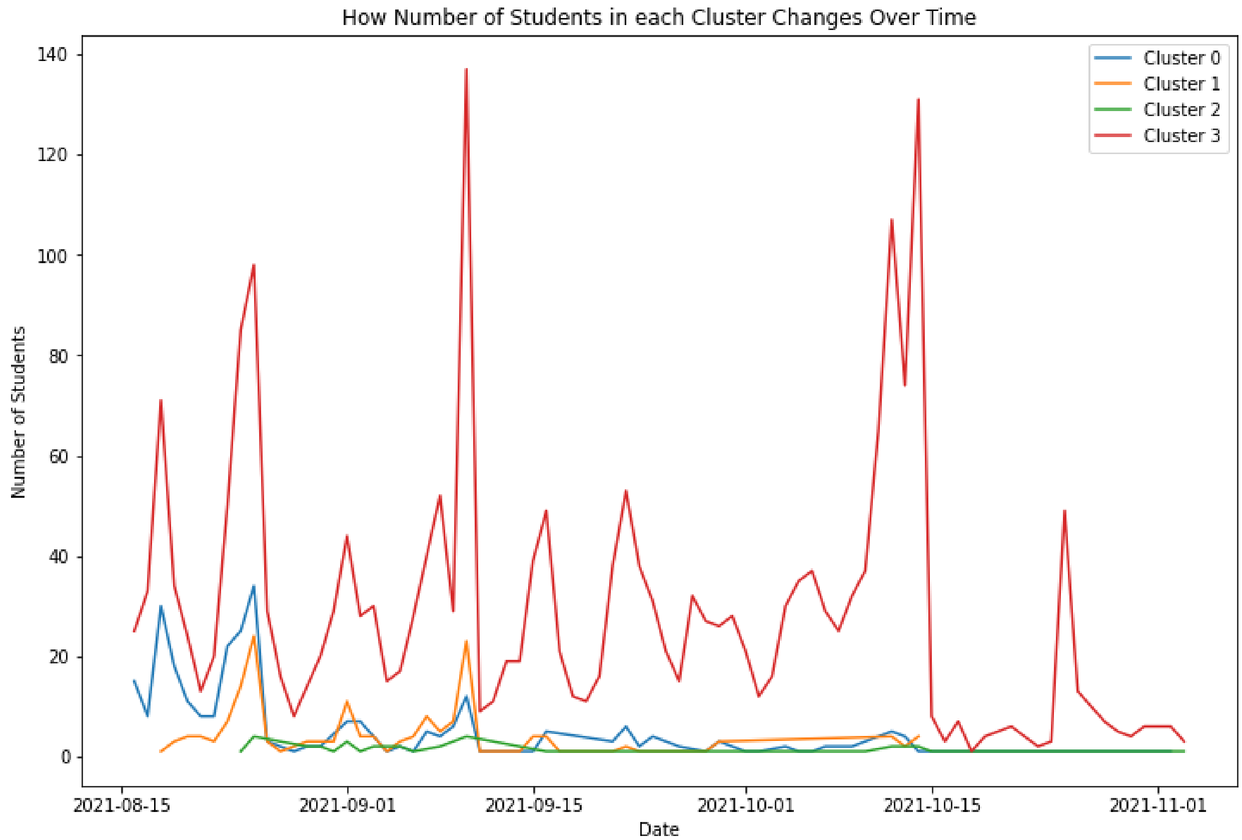


Figure 4.21: Evolution of Students in Clusters During the Online Course

4.4.3 Post Study Results

This section describes the findings from post studies that were carried out on participants from the experimental system. The post studies comprised of a structured survey and a semi-structured interview.

4.4.3.1 Post Study Survey

The post-study structured survey was used to measure students' satisfaction and perception of the benefits of the EDM interventions. The purpose of this survey was to establish students' perception of the usefulness of the intervention messages that were provided by the LMS during the course.

Students who completed the course were voluntarily invited to participate in the survey and give their feedback on the usefulness of the EDM tool in providing the SRL intervention messages.

(i) Participants' Demographic Features

The participants in this post-study survey were students who had completed the online course using the experimental system. A Google form link was submitted to all the participants and a total of 141 students responded. 74.5% were male and 25.5% were female. 90.1% were students pursuing degree programs while 9.9% were pursuing diploma programs. The low percentage of diploma participants was expected since the study involved university students of which the majority pursue degree programs. 72.3% indicated that they had previous experience in online/blended learning while 27.7% had no online/blended learning experience.

From Table 4.26, it can be observed that students very often communicated to their colleagues and instructors using social media platforms such as WhatsApp and Facebook. The students also indicated that they often use phone calls, emails, and text messages.

Table 4.26: Tools Students Utilized to Communicate to Peers and Instructors (N=141)

Tool	Very Often (%)	Often (%)	Sometimes (%)	Rarely (%)	Never (%)
Email	24.82	27.66	24.11	16.32	7.09
Text Messages (SMS)	26.24	22.70	19.15	17.73	14.18
Social Media (WhatsApp/Facebook)	69.50	19.15	9.22	0.00	2.13
Phone Calls	34.04	26.24	18.44	7.80	13.48

The participants who did not have prior experience in data science or machine learning were 74.5% while 25.5% had indicated that they had experience in data science or machine learning.

Table 4.27 presents the summary of the demographic information for the participants.

Table 4.27: Summary of Demographic Features for Structured Survey Participants(N=141)

	Demographic features	N	%
1	**Gender**		
	Female	36	25.50
	Male	105	74.50
2	**Level of Study**		
	Degree	26	90.10
	Diploma	406	9.90
3	**Previous experience in online/blended learning**		
	Yes	102	72.30
	No	39	27.70
4	**Previous experience in data science or machine learning**		
	Yes	36	25.50
	No	105	74.50

(ii) Structured Survey Results

a. Students' Experience on the Benefit of Interventions Messages

The results of students' experiences on the benefits of the intervention messages provided by the EDM plugin are presented in Table 4.28. The results indicate that the intervention messages helped the students enhance their online learning behavior. For example, 76.59% of students indicate that their individual needs were met by providing individualized intervention messages compared to 23.41% who were either neutral or disagreed that their individual needs were not met. 73.04% of students indicated that with the help of the intervention messages they were able to enhance their engagement levels while 87.94% of students indicated that by providing the summary of learning activities, the LMS helped them improve on the areas they were not doing well. As indicated, 73.05 % of students said that the intervention messages provided by the LMS helped them improve poor learning habits.

Table 4.28 presents the responses on students' perceived benefits of the intervention messages.

Table 4.28: Students' Responses on the Benefits of EDM Interventions (N=141)

Item	Scale	N	%
The LMS provided a platform that made me feel that my individual needs were met by providing individualized intervention messages	Strongly Agree	62	43.97
	Agree	46	32.62
	Neutral	28	19.86
	Disagree	3	2.13
	Strongly Disagree	2	1.42
The learning intervention messages provided by the LMS helped me enhance my engagement level	Strongly Agree	58	41.13
	Agree	45	31.91
	Neutral	32	22.70
	Disagree	4	2.84
	Strongly Disagree	2	1.42
By providing the summary of learning activities, the LMS helped me improve on the areas I was not doing well	Strongly Agree	76	53.90
	Agree	48	34.04
	Neutral	16	11.35
	Disagree	1	0.71
	Strongly Disagree	0	0.00
The intervention messages provided by the LMS helped me improve on poor learning habits	Strongly Agree	55	39.01
	Agree	48	34.04
	Neutral	29	20.57
	Disagree	7	4.96
	Strongly Disagree	2	1.42

b. Students' Responses on how LMS enabled them to Utilize SRL Strategies

In Table 4:29, we demonstrate students' responses on how the LMS (which had the EDM plugin installed) helped them utilize various SRL strategies. For instance, 83.69% indicated that the LMS helped them enhance time management, 88.65% indicated that they improved self-monitoring skills and 90.78 % indicated that they were able to develop self-evaluation skills. On help-seeking, 63.83% enhanced help-seeking from among student colleagues while 68.80% enhanced help-seeking from instructors. 86.52 % of students further indicated that they were able to improve effort regulation skills in terms of knowing the areas where they needed to commit more time and effort. Finally, 87.94% of students were able to plan and organize their course learning activities and schedules through the support offered by the EDM plugin via the LMS.

Table 4.29: Students's Responses on how the LMS helped them Utilize SRL Strategies (N=141)

SRL Strategy	Scale	N	%
Manage my time well while studying online	Strongly Agree	52	36.88
	Agree	66	46.81
	Neutral	23	16.31
	Disagree	0	0.00
	Strongly Disagree	0	0.00
Monitor my online learning strategies	Strongly Agree	61	43.26
	Agree	64	45.39
	Neutral	14	9.93
	Disagree	2	1.42
	Strongly Disagree	0	0.00
Evaluate myself concerning my progress in learning activities to achieve course goals	Strongly Agree	71	50.35
	Agree	57	40.43
	Neutral	13	9.22
	Disagree	0	0.00
	Strongly Disagree	0	0.00
Seek help from my colleagues	Strongly Agree	38	26.95
	Agree	52	36.88
	Neutral	42	29.79
	Disagree	7	4.96
	Strongly Disagree	2	1.42
Seek help from the course instructor	Strongly Agree	32	22.70
	Agree	65	46.10

	Neutral	39	27.66
	Disagree	4	2.84
	Strongly Disagree	1	0.70
Effort management in terms of knowing the areas where I needed to commit more time and effort	Strongly Agree	60	42.55
	Agree	62	43.97
	Neutral	19	13.48
	Disagree	0	0.00
	Strongly Disagree	0	0.00
To plan and organize my course learning activities and schedules	Strongly Agree	64	45.39
	Agree	60	42.55
	Neutral	16	11.35
	Disagree	1	0.71
	Strongly Disagree	0	0.00

c. Students' Responses on the Support Received during the Online Course

The study also captured students' responses to the support received during the online course as presented in Table 4.30. Most of the students indicated that they were satisfied with the level of support they received and the timely instructional feedback in terms of the intervention messages at 85.00% and 77.05% respectively. From the results, 56.73 % of students indicated that the email notifications they received helped them track the intervention messages easily while 75.18% said that they were satisfied with the regular reminders that they received on how to utilize the LMS features to improve their learning engagement.

Table 4.30: Students' Responses on the Support Received during the Online Course (N=141)

SRL Strategy	Scale	N	%
I was satisfied with the level of support I received through the LMS	Strongly Agree	58	41.13
	Agree	62	43.97
	Neutral	20	14.18
	Disagree	1	0.72
	Strongly Disagree	0	0.00
The instructional feedback in terms of the intervention messages I received was timely	Strongly Agree	46	32.62
	Agree	57	40.43
	Neutral	30	21.28
	Disagree	6	4.25
	Strongly Disagree	2	1.42
The email notifications that I received helped me track the intervention messages easily	Strongly Agree	33	23.40
	Agree	47	33.33
	Neutral	38	26.95
	Disagree	14	9.93
	Strongly Disagree	8	6.39
I was satisfied that my individual learning needs were catered for during the online course	Strongly Agree	53	37.59
	Agree	54	38.30
	Neutral	31	21.99
	Disagree	3	2.12
	Strongly Disagree	0	0.00
I was satisfied with the regular reminders that I received on how to utilize the LMS features to improve my learning engagement	Strongly Agree	53	37.59
	Agree	53	37.59
	Neutral	29	20.57
	Disagree	2	1.42
	Strongly Disagree	4	2.83

d. Students' Best Learning Experience during the Online Course

From an open-ended question, students were asked to give their best learning experiences during the online course.

Table 4.31 presents some of the highly cited experiences by students.

Table 4.31: Students' best Learning Experience during the Online Course

Thematic area	Cited Responses
Online learning support	"I was satisfied with the level of support I received through the LMS", "I was satisfied with the regular reminders that I received on how to utilize the LMS features to improve my learner engagement", "I was satisfied that my individual learning needs were catered for during the online course", "The motivational messages", "Regular updates from the instructor via various platforms".
Learner engagement	"Engagement with the quizzes and assignments", "Interacting with the online learning materials on LMS", "The hands-on interaction was superb", "Engagements were timely, friendly, and helpful", "I really enjoyed the interaction with the materials provided i.e. YouTube videos", "Interacting with the materials were very resourceful and I got to learn much than I expected", "The online interaction was good and knowledge well disseminated", "group forums".
Enhancement of SRL strategies	"Getting to change my self-driven behavior of pushing myself to study", "peer interaction", "The best online experience I had is interacting with other students", "Having interactions with my peers and online instructor", "asking a question as is brainstormed by people from different universities", "Ability to go back and recall previous lecturer briefings", "self-paced learning experience came in handy", "I was able to learn out of

	my busy schedule”, “development of ‘self-driven’ attitude throughout the course”.
General sentiments	“It was one of the best online learning system I have ever had”, “It was the best experience”, “user friendly”, “the very best”, “I was able to understand LMS”, “getting to learn more all in one place”, “I enjoyed to have completed machine learning in such fashionable way”, “availability of the materials throughout”, “organization of topics that made it easy to follow”, “was awesome”.

e. Students' worst Learning Experience during the Online Course

From an open-ended question, students were asked to give their worst learning experiences during the online course. Table 4.32 presents some of the highly cited experiences by students.

Table 4.32: Students' worst Learning Experience during the Online Course

Thematic area	Highly cited Responses
Internet access	“lack of internet to sufficiently access the course”, “lack of bundles and power blackout”, “network failure”, “power blackout”, “data scrambling”, “weak network connectivity”, “poor quality network”, “power outage”, “poor internet connectivity”, and “inadequate internet”.
Physical interaction	“lack of interaction with colleagues”, “distance learning whereby it was difficult to do team work with my peers”, “learning online was not being able to interact with my peers and course instructor”, “lack of peer-to-peer study facilitation”, “lack of one-on-one engagement with my instructor”, and “lack of support from colleagues”.
Intervention messages	“motivation messages were boring”, “messages from LMS sometimes were too much”, and “too many reminders”
General sentiments	“none”, “I did not have bad experience during the entire course”, “there is nothing to complain about”, “time factor”, “there was no worst experience”, “no detailed explanation for beginners”, and “deadlines for the tests and poor time management from my side”

f. Additional Features and Areas of Improvement on the LMS Interventions

Through an open-ended question, we sought to establish the additional features and areas of improvement in LMS intervention messages. Table 4.33 highlights some of the frequently mentioned responses.

Table 4.33: Suggested Areas of Improvement on the LMS Intervetions

Thematic area	Sample Responses
Intervention messages	“Ensure that email server can send email notifications to all students”, “do not popup notifications full screen, perhaps put them on the side”, “and reduce the number of notifications”, and “the comments that the LMS was providing were encouraging”
Communication	“include a real-time chat system into the LMS that enables you to directly ask your instructor questions”
Internet access	“Improve on internet provision”, “make materials available offline”, and “provision of free internet bundles”.
General suggestions	“I would suggest that you include more assignments and quizzes to keep us on toes on areas we need to put extra effort”, “more learning materials”, “more practical assignments”, “more online assignments”, “increase the number of lessons attended by instructors in a week”, “more content”, and “more quizzes”

The results demonstrate that students had the best experience. Furthermore, students indicated that the frequency of intervention messages presented on the LMS was disturbing and recommend that the frequency be reduced.

4.4.3.2 Post Study Interview

A semi-structured interview was carried out with students who had used the experimental system during the online course. The study aimed to qualitatively establish the students' perception of the benefits of the interventions messages that were sent out to them via LMS during the online course and obtain any suggested ways on how the interventions can be presented to the learners. To achieve this, a set of similar questions were asked to each participant. Moreover, the respondents were allowed to give open-ended responses in areas they deemed relevant as per their online learning experience and interaction with the intervention messages.

(i) Participants

A voluntary email invite was sent out to the 122 students who had completed the Data Science with Python course. A total of 20 students accepted to take participate in the interview. The interview was carried out using the Google Meet link and recorded and the audio was transcribed using the otter.ai application.

(ii) Semi-Structured Interview Results

Thematic content analysis was applied to the interview transcript that had been converted into text. This process involved aligning the data into themes that were common during the interview. The findings from the analysis were categorized into four thematic areas;

a) Students' feedback on interventions messages
b) Suggested additional features to the EDM plugins
c) Other approaches to delivering intervention messages
d) Students' general online learning experience

Table 4.34 presents the identified thematic areas and cited responses from the participants.

Table 4.34: Thematic Areas identified from post-study Interview and Sampled responses

Thematic area	Cited Responses
Students' feedback on intervention messages	"*They (intervention messages) really helped me a lot…I got to a certain point that I felt like giving up, I almost gave up actually but when the intervention messages came up. And so that I was I was doing great with engaging the LMS and that gave me the encouragement I needed to move on. And keep trying to learn something new in the course*."
	"*…the popup messages that that used to provoke us. They could add more sense to people losing hope… I was new to LMS the messages encouraged me that wherever I could move away maybe I could skip maybe one lesson or two. When I could log in, the message would welcome me…and that encouragement really helped me to push on and even learn more until I could come to the end of the course*."
	"*Based on my experience, I can say the intervention messages actually helped me in my part, not through the emails because I had left you my emails. But the one I was like the one that was popping up in the in the LMS Yeah, get maybe like me, Lesson 8 was kind of hectic like, I was not getting some of the concepts but through the messages, there was kind of motivation that kept me moving forward, actually*"
	"*I think the popup was quite fine, but at some point, it used to be kind of nagging, like they kept popping up each time. So I think maybe, if possible, you can have it maybe pop up like one moment I accessed the LMS it pops up and then later on and acquisitive allied resources it doesn't pop up but we'll be better*"
	"*…concerning the online intervention, the messages, they were okay there was a time that they were a bit many messages at once on the site*

	but along the way, you reduced the number and that was okay. Because for someone like me who was a beginner who knew nothing about data science and everything. I found it hard at the beginning but after a while after engaging with the LMS those messages encouraged me and made me to keep going on to keep studying and keep in trying to know more…for me that was the best experience."
Additional features	Addition of collaborative learning through group projects "*Is it possible to introduce the project in a group and another one that is done individually? Group one I think will help me learn a lot about collaboration and also teamwork."* Integration of third-party applications such as Slack that can *allow group collaboration and sharing of documents by allowing students "interact with each other as a group, or individually or as an individual share document. That's when a person for example, without their phone, they still can have access to communication."* "*…if you can find a way of determining the students who have dropped out of the course so that you can talk to them and know why they have dropped out or is there anything that has made them drop out of the course*."
Other approaches to delivering intervention messages	The use of text messages can help reach "*the ones that stay offline without engaging the materials on the LMS… Maybe if someone doesn't engage the materials for a week or something we should find a way to reach them outside the LMS environment so that maybe you can be one of them*." "*I think also sending text messages will be better. Because many people do not have access to let's say WiFi that can enable them to be online throughout but they can read their emails*"

	Identifying active learners and rewarding them. "*One of the other ways of maybe keeping students involved you will have come up with something maybe to do with the champion of the week*"
General learning experience	"*Actually, the course was good. And I just want to thumbs up the LMS team unity we put it up great work*."
	"*You know, some of us actually began from scratch and for us to gain for us to begin from scratch and in 10 weeks you have something in mind is a big achievement*."

4.5 Discussion on Research Findings

The discussion of the results is presented according to the following research questions and hypotheses that were being tested in this research:

- RQ1: What is the current status of utilization of LMS features in promoting SRL strategies in online courses?
- RQ2: How effective is the EDM model in promoting SRL on LMS?
- RQ3: What is students' perceived usefulness of the EDM model in promoting SRL on LMS?

Research Question 1 (RQ1)

What is the current status of utilization of LMS features in promoting SRL strategies in online courses?

First, the study investigated the utilization of LMS features in enhancing SRL. The findings indicate that chats, forums, quizzes, and messages are among the features that are commonly used by learners. The wikis, blogs, and workshops are least used by learners. Likewise in a related study carried out by (Back et al., 2016), it can be observed that LMS features such as wikis and blogs are least used and therefore required by the learners. From another related study, it is observed that the wikis acceptance levels by students are low (Yilmaz et al., 2017). From the literature, all the features considered in the present study are significant in improving students' learning experience. Janson et al. (2017) investigated the factors that influence learning processes and success in LMS. The study indicated that students perceive learner support, interactivity with peer learners and instructors, and Task-Technology Fit (TTF), which refers to the LMS features (forums, quizzes, assignments) that enable learners to carry out learning activities such as communication and interaction with peers and instructors, positively influence the use of LMS. From this study, it can also be noted that communications tools are underutilized as emails appear to be preferred amongst students. In a related study, Back et al. (2016) investigated how medical students utilized LMS for their learning and the challenges they faced. The study revealed that the students mainly communicated via emails and Facebook. The most popular tools the students used in their learning were slides, video, and digital texts. Interestingly, it was also observed that in using LMS, there is a lack of interaction. The students also reported that there is a lack of personalized interaction between them and instructors and this required them to make much effort in learning. This study concurs with related studies where students suggested the incorporation of telephone messages and social media platforms as integrated channels of communication (Back et al., 2016). Social platforms such as Facebook if integrated into online learning environments could be useful in group social collaborative learning by students which could potentially help students develop socially shared regulation skills (Yilmaz et al., 2017).

Secondly, from the findings in the present study, it can also be observed that the LMS inbuilt features such as discussion forums, chats, quizzes, emails, and wikis that are supposed to stimulate the growth of SRL skills are underutilized and therefore not likely to promote the development of SRL. The implication is that learners are not benefiting from the features due to a lack of SRL skills and may require interventions to enable them to apply the skills. It is important to observe that the inbuilt LMS features if well utilized by learners and instructors could lead to satisfaction and success in online learning (Vovides et al., 2007). According to Oliveira, Cunha, and Nakayama (2016) technology should be an enabler that allows for high-level interaction, collaboration, and communication among students and instructors especially when learning takes place through LMS. The findings in this study imply that for learners to develop and grow their SRL skills, they require a learning environment that supports them in planning and monitoring their learning. As Palomino et al. (2014) assert, there is a need to integrate the benefits of intelligent tutoring systems into LMS to experience aggregated advantages of both. Additionally, to curb this problem of underutilization, there is also a need for university students and instructors especially those who participate in online learning and teaching be trained on the importance of Self-regulated learning (Nunez et. al.,2017). Furthermore, providing SRL interventions benefits students to plan, monitor, and reflect on the learning practices by providing relevant guidelines and hints to improve their self-regulatory skills, especially in online learning environments (Viberg et al., 2020). According to Kim et al. (2019), online learning environments need to be designed in a way that supports and deepens students' engagement levels through integration with intelligent tools. The intelligent tools can help infer learning behaviors from log data and then provide targeted interventions for each learner.

Lastly, this study sought to know students' perceived challenges encountered during online learning. The student responses indicate a lack of real-time and individualized feedback especially to guide students in their learning habits and performance in various course assessments such as continuous assessment tests and quizzes. The results also reveal that a lack of internet connectivity is a hindrance to active and successful online learning. The qualitative findings concur with the findings in previous studies where students pursuing open and distance learning experience poor internet connectivity (Muuro & Kihoro, 2017). From literature, it can be observed that only those LMS that have incorporated hypermedia and provide personalized feedback for learners can lead to increased self-regulated learning and improved motivation. The feedback helps the learner to be able to utilize LMS resources (Sáiz-Manzanares et al., 2019). Similarly, according to Fetzner (2013), when students are left alone without guidance and support, they feel left behind and are likely to drop out or fail online courses. The challenges experienced by learners can also be attributed to the underutilization of various SRL strategies in online courses. It will be interesting to consider the provision of real-time and frequent feedback on how students utilize LMS features to support various learning activities. Real-time feedback to students helps them not only become aware of their learning habits but also know how to improve their learning speed so that they are not left behind. When students feel they are left behind they may feel dissatisfied and demotivated which has also been attributed to unsuccessful online learning (Fetzner, 2013). Hashemyolia et al. (2014) also argue that although most online learning environments offer a variety of tools and features that are intended to enhance engagement levels, the actual use by the students depends on their levels of motivation.

Finally, the pre-study findings helped identify the features that had a direct influence on the design of the experimental system. The pre-study carried out on students from selected universities in Kenya, who also served as intended users, was used to investigate whether the current LMS supported learners in improving SRL skills. The survey also helped identify the learning activities that students performed most and their expected needs. The study captured the requirements specification and system design that guided the system development.

Table 4.35 presents the key findings and how they guided the system development.

Table 4.35 The Pre-Study Findings that Guided System Development

Problem	Solution
Lack of individualized feedback	Designed an algorithm that assigns students into groups based on their learning behaviors. Similar groups of learners were then used to offer targeted feedback based on learner behaviors as captured by learning activities engaged on
Lack of instructional support	Included metacognitive prompts to guide students in their study
Minimal involvement in learning activities based on chats, discussion forums, and quizzes	Collected data on the use of these features and used the analysis to support learners through targeted interventions

Research Question 2 (RQ2)

How effective is the model in promoting SRL on LMS?

For this research question the following hypotheses were formulated:

1. Null Hypothesis: There is no significant difference between SRL strategies utilization by students in EDM-based intervention and instructor-based intervention groups.
2. Alternative Hypothesis: There is a significant difference between SRL strategies utilization by students in EDM-based intervention and instructor-based intervention groups.

From the Mann-Whitney U test results presented in Table 4.20, it can be observed that there was a significant difference between the mean for effort regulation SRL strategy. There was no significant difference between the means for control of learning belief, rehearsal, time, and study environment and help-seeking among the students in the control and experimental groups. Therefore, for the effort regulation SRL strategy, we reject the null hypothesis and fail to reject the alternative hypothesis. However, for control of learning belief, rehearsal, time and study environment, and help-seeking, the alternative hypothesis is rejected and we fail to reject the null hypothesis.

Based on the self-report posttest data obtained using the MSLQ scale, the EDM plugin was only able to enhance students' effort regulation skills. However, applying independent t-test analysis on LMS log data as presented in Table 4.24, there was a significant difference between the means of time on task, rehearsal, time management, and help-seeking between the experimental and control group. The null hypothesis is therefore rejected and the alternative hypothesis is not rejected. Therefore, there is a significant difference between SRL strategies utilization by students in EDM-based intervention and instructor-based intervention groups. The results provide evidence that the EDM plugin was effective in promoting SRL strategies in the learning management system. It is interesting however to note that what students reported using the self-report SRL measurement tool, the MSLQ scale, differs from the actual observations and measurements from the LMS log data. This provides empirical evidence that the use of EDM techniques in measuring SRL provides an unobtrusive approach to measuring SRL skills for students without capturing the bias and misreporting by students when using the self-report tools. Consequently, analysis of LMS data for inferences in SRL strategies provides the best alternative to identifying students' SRL skills compared to self-report tools. This finding is in agreement with previous studies that show that self-report tools are biased and only capture SRL strategies based on the way students perceive themselves and that self-report tools that were designed for use in face-to-face classrooms may not be effective in online learning environments such as LMS (Araka et al., 2020; Roll & Winne, 2015; Winne & Baker, 2013).

The findings in this research reveal that EDM techniques are effective in promoting self-regulated learning strategies such as cognitive and behavioral strategies. Students who received intervention messages may have leveraged the suggestions and hints to engage more in learning materials and participate in learning activities such as discussion forums and quizzes that lead to stimulation and utilization of cognitive and behavioral strategies that were under study. Notably, these findings align with recent related studies. For instance, a real-time chatting system for enhancing help-seeking has also been used to promote self-regulated learning in higher education for online and blended learning (Broadbent et al., 2020).

Previous studies also indicate that offering intervention messages to online learners enhances self-regulatory skills such as self-efficacy (Müller & Seufert, 2018). The intervention messages comprised of individualized suggestions and hints that were designed to influence the learner to engage more in learning materials and activities to develop the underlying SRL skills. The messages, therefore, served as real-time feedback on learners' study behavior. The feedback messages were also presented to students via the student dashboard which ensured those who disabled the pop-up messages were still able to access the messages.

Moreover, email notifications were also provided to support the students who rarely accessed LMS to read the messages sent within the LMS. The messages stored on the student's dashboard were implemented using an approach that compelled those who may have disabled the pop-up widget to access the messages. Since online learners are not controlled and restrained to study within given schedules tracking learner interactions within the LMS and providing intervention messages to them enhances their level of engagement with learning materials and communication with each other and ensures that students are placed at the center of the learning process. As presented in Figures 4.16 and 4.17, the students in the experimental group who were enrolled in the experimental system engaged more compared to their counterparts in the control group respectively. When students are provided with learning analytics summary reports via dashboards, self-regulated learning skills such as self-monitoring and evaluation are enhanced as well as learner engagement which leads to increased academic performance (Yoon et al., 2021). The results of this study reveal that offering feedback based on students' learning behavior enhances learners' ability to utilize LMS features and enhances students' self-regulatory skills.

The findings in this research align with current research. For example, Chou & Zou (2020) developed and implemented a tool that allowed students to receive external feedback that enhanced their internal SRL processes, monitoring, and performance. However, Chou & Zou (2020) suggest that the external feedback may not work well for all the students and therefore propose that such students could receive further support that may allow negation between the students and system (integrated with SRL promoting tools) and allow co-regulation with other students. Sáiz-Manzanares et al. (2019) argue that offering individualized feedback based on analytics from intelligent agents motivates students to engage more especially in self-paced learning. Furthermore, Sáiz-Manzanares and his colleagues argue that EDM techniques provide possibilities of providing targeted interventions hence enabling students to monitor their learning based on actual learning behavioral feedback. The outcomes in this research also agree with related research conducted by Karaoglan-Yilmaz and Yilmaz (2021), which investigated the effect of providing personalized metacognitive feedback on students' engagement in online learning. The feedback, which was comprised of analytic reports and metacognitive messages, was provided to the learners based on the recommendations from learning analytics. The findings reveal that the engagement level for students who received the interventions increased significantly compared to those who did not receive the interventions. Moreover, Viberg et al. (2020) argue that providing intervention messages helps students in planning, monitoring, and reflecting on learning habits in online learning environments where instructor support is limited. Related research shows that students with poor internal SRL processes often lead to poor monitoring and performance and a lack of motivation to learn. External tools that provide external feedback and support to enhance students' SRL skills are required. Such tools help learners to monitor their learning and enhance their motivation by allowing students to reflect on their learning behaviors and performance (Chou & Zou, 2020).

Research Question 3 (RQ3)

What is students' perceived usefulness of the EDM model in promoting SRL on LMS?

The post-study survey and interview sought to qualitatively establish students' perceived benefits of the EDM plugin in providing intervention messages and learning progress on LMS. First, the findings indicate that the intervention messages helped the learners to reflect and improve on their undesirable habits and enhance their engagement levels in the LMS courses. The findings further reveal that the SRL intervention messages helped students improve on utilizing various SRL strategies including time management, self-monitoring, help-seeking, and effort regulation. Through the intervention messages, the learners were able to identify the areas in the course and the academic activities to which they needed to commit more time and effort. In a related study, students indicated that they require tools that help them to plan, monitor, schedule, and regulate their studies by use of progress tracking tools as well as receiving timely feedback on their learning behavior (Alasalmi, 2021).

Second, from the students' responses, it can be observed that most of the students were satisfied with the real-time instructional feedback that they received via LMS as well as the regular reminders on how to utilize LMS features to improve their learning. From the sampled student responses, it can be revealed that the EDM plugin helped students to develop the self-learning habit of being in control of their learning. For instance, when asked to record their best online learning experience, students indicated; "*getting to change my self-driven behavior of pushing myself to study*", "*self-paced learning experience came in handy*" and "*development of self-driven attitude throughout the course.*"

Lastly, despite the increased level of support that led to students' satisfaction with the online course, the students faced some challenges including lack of internet access and lack of one-on-one interactions with peers and course instructors. Previous studies also identify internet access as a major hindrance to online learning especially in developing countries (Araka et al., 2021; Muuro & Kihoro, 2017). On the areas of improvement, students indicated that integrating the LMS with third-party tools such as "slack" could enhance collaboration in online learning where students can share knowledge, challenges, and experiences among themselves. Moreover, students indicated that the use of text messages (SMS) instead of email notifications and the internal LMS reminders could help reach those who are often offline due to lack of internet access.

The findings from the post-study survey and interview confirm that students were satisfied with the intervention messages that helped them stimulate SRL skills. For instance, a student reported that the best learning experience was "*getting to change my self-driven behavior of pushing myself to study*." The students further indicate that to enhance the online learning experience, there is a need to integrate the plugin with third-party tools like "*Slack"* that can enhance collaborative learning. Moreover, it will be more effective to provide students with offline text messages to curb the challenges encountered by those unable to access email and LMS due to lack of internet access.

Finally, supporting students' SRL strategies likely to impact completion and retention rates. For example, in the current study, the students who received SRL interventions and completed the online course were more compared to those who did not receive the interventions (experimental group n=122; control group n=109) as presented in Figure 4.15. Moreover, the SRL interventions enabled the students who were giving up got reenergized and motivated to continue with the online course. For instance, a student in the post-study survey reported "*those messages encouraged me and made me keep going on to keep studying and keep in trying to know more*". Another student reported "*…I was new to LMS the messages encouraged me wherever I could move away from maybe I could skip maybe one lesson or two. When I could log in, the message would welcome me…and that encouragement really helped me to push on and even learn more until I could come to the end of the course*".

4.6 Summary

True experiment research was used to establish the effectiveness of the EDM model (which was designed and implemented in this study) in promoting SRL through sampled university students who enrolled in data science with python course for 12 weeks. The course was delivered through Moodle LMS. Moodle LMS stores log data that captures students' interactions with collaborative and communications features such as forums, chats, quizzes, assignments, and learning resources which, if well utilized, enables learners to be active in learning processes. The log data was collected, preprocessed, and analyzed using EDM algorithms namely agglomerative hierarchical clustering and logistic regression algorithms. The EDM algorithms were utilized to identify students' cognitive and learning behaviors and deliver SRL interventions through prompts and real-time feedback on the visualized dashboard. The clustering algorithm was used to identify four clusters of students: poor self-regulators, moderate self-regulators, good self-regulators, and exemplary self-regulators. The clustering process was determined by the frequency of views on resources and learning content and students' interactions with one another. The classification algorithm was used to place students into the right cluster progressively in real-time; each time they accessed the online course. The SRL interventions were delivered to students based on their cluster behaviors and engagement levels. The EDM-based interventions were presented to students enrolled on the experimental system (with the EDM plugin installed) via a visualized dashboard on Moodle LMS and email notifications. The students in the control group who were enrolled on the non-experimental system (without the EDM plugin installed) received the usual instructor-led interventions. Quantitative data from the study was collected using motivated strategies for learning questionnaires for pretest and posttest. Time series data from Moodle LMS was also obtained for analysis and inference of the self-regulated learning strategies. A post-study that involved a semi-structured survey and semi-structured interview was conducted to establish students' perception of the benefits of the EDM-based interventions. The independent t-test on the pretest and post-test data and the LMS log data reveal that the provision of EDM-based interventions had a significant impact on students' utilization of self-regulatory skills. The findings reveal that the EDM plugin was effective in providing targeted intervention messages to students and supporting the growth of SRL strategies. Students' perception of the usefulness of the support they received from the LMS where the plugin was installed to promote SRL should be addressed as suggested by the learners in the post-study survey and interview. In conclusion, this research has provided insights into identifying disengaged students and providing scaffolds to motivate them and get them back to track in online learning environments.

CHAPTER 5
CONCLUSION AND RECOMMENDATION

5.1 Introduction

This research was conducted to examine the promotion of SRL using educational data mining techniques. First, a systematic literature review was conducted to investigate the current methods being used to measure and promote SRL in online learning environments. Second, a pre-study was conducted to establish the current status of the utilization of LMS features in promoting SRL in online courses. Third, an empirical evaluation of educational data mining techniques for profiling SRL in online learning environments was carried out. This study led to the design and implementation of an EDM plugin that provides intervention on LMS to promote SRL. Lastly, the effect of the interventions on students' SRL strategies was investigated using a true experimental research design on an actual LMS.

This chapter presents the conclusions made from the study, research contributions, and future work that carried be carried out.

5.2 Conclusion

Open and distance learning mode has continued to attract a large number of students who find it convenient and affordable to study online. The main limitation of online learning has been the lack of students' support compared to their counterparts in face-to-face classrooms. Particularly, online learners are supposed to be supported so that they develop self-regulatory skills and utilize them during the learning process. Although machine learning has been applied to educational data to understand learners and learning environments, little evidence exists on the effect of providing actionable feedback to students on learner SRL skills and academic performance. Specifically, the focus of this research was collecting and analyzing LMS log data in real-time using educational data mining to measure and promote SRL through automated and targeted intervention messages. The findings from this research reveal that offering actionable metacognitive feedback through visualized dashboards and pop-up widgets helps students to participate more and get involved in online academic activities leading to the development of SRL skills.

Identifying clusters from LMS log data provides an opportunity for the design of individualized and targeted interventions geared towards enhancing students' self-regulated learning skills. This research explored how to identify and promote self-regulated learning strategies from trace data in learning management systems. The analyzed data was used to understand online learning behaviors for students (where learners are) and provide them with intervention messages to activate and stimulate their self-regulatory skills (to take them to where they are supposed to be). The study explored how to promote SRL strategies identified from two SRL categories as defined by Pintrich's SRL theoretical model; motivation and learning beliefs (behaviors).

The intelligent tool referred to as the EDM plugin was designed, developed, implemented, and evaluated. The tool provides clustering and classification functionality that allows new instances of students accessing a course on LMS to be placed in clusters and start receiving SRL intervention messages. The interventions are presented to students via pop-up widget and email notifications. Furthermore, students who ignore the messages on email or within the pop-up widget can still access the interventions through a message container on the students' dashboard. The EDM plugin was implemented on Moodle LMS where the effect of intervention messages on students' levels of engagement and SRL learning behaviors were measured. This was achieved by tracking and collecting students' traces into the EDM model that was integrated with clustering and classification algorithms that placed students into relevant clusters and provided intervention messages based on their learning behavior. The traces were captured and stored by the Moodle core DBMS. Moreover, the study examined the relationship between the students' SRL skills and academic performance. The aim was to establish where students are and present to them metacognitive messages that could enable them to pause learning and reflect on the learning behavior. The metacognitive messages were aimed at stimulating the learner's SRL skills and therefore enabled the measurement and promotion of SRL at the same time (Araka et al., 2020). The SRL measurement was depicted on the placement of students into clusters and promotion through the provision of intervention messages presented to the learner via a visualized dashboard, pop-up window, and email notification to students.

This study offers insights into how instructor-learner interaction support can be enhanced through the use of educational data mining. Although researchers argue that the risk involved in this direction is losing social interaction (Seo et al., 2021), artificial intelligence at least, for now, has proven to be the preferred approach that enables the provision of support to a large number of online students despite having a limited number of instructors for online facilitation since the findings indicate that promoting SRL in online learning environments leads to improved learner engagements, SRL skills and learning outcomes in terms of academic performance. The adoption and use of the EDM plugin in higher institutions of learning, where learning now is majorly online, is likely to lead to more students engaging in academic learning activities, and better learning outcomes as a result of their SRL skills being enhanced. The novel approach allows the promotion of SRL on LMS and demonstrates that it is possible to overcome the challenge of lack of instructor support due to the limited number of instructors who facilitate online courses in higher institutions of earning amidst growing student enrollments for online learning. This study has demonstrated that providing external support through the use of dashboard analytics and real-time feedback leads to the growth of students' SRL skills as proposed by (Karaoglan-Yilmaz & Yilmaz, 2020; Karaoğlan-Yılmaz et al., 2018). Moreover, this study has proven that it is possible to achieve the recommendation for the third wave of SRL measurement and promotion by developing and implementing intelligent pedagogical agents that can measure and promote SRL on LMS in concurrence (Panadero et al., 2016).

Generally, this study has demonstrated that students on LMS can be supported to enhance their SRL skills using educational data mining techniques. The EDM plugin that was developed, implemented, and evaluated in this study was able to deliver real-time SRL interventions to online students. In this study, we implemented an EDM approach to promote SRL by developing an EDM plugin that was guided by two theories; constructivism and cognitive theories of learning. The EDM plugin that was integrated into Moodle LMS mined log data from Moodle core database and provided SRL interventions to learners in form of the visualized dashboard, pop-up widget, and email notifications in real-time. The experiment was conducted by involving university students in online data science with python course that was delivered via Moodle LMS. The findings from this study reveal that LMS designed and reinforced by the theories of constructivism and cognition and integrated in form of an EDM plugin not only has the potential of enabling students to learn actively but also promotes students' SRL skills hence improving learner motivation, satisfaction, and completion rates in online courses. The findings also indicate that EDM algorithms embedded within LMS improve self-regulated learning skills such as time on task, rehearsal, time management, and help-seeking. This study reveals that self-regulated learning skills for online learners can be enhanced by integrating EDM algorithms on LMS. This ensures that actual LMS log data can be inferred and used to understand learner behaviors and therefore providing an effective way of proposing targeted interventions to the students.

The results in this study agree with previous related research. For example, Susnjak et al. (2022) developed a learning analytics dashboard that utilizes descriptive analytics and machine learning algorithms to provide predictive and prescriptive insights to learners. The study proposes that providing automated interventions to learners via the dashboard and investigating their effectiveness in supporting personalized can be implemented in the future. Khiat and Vogel (2022) developed and evaluated a self-regulated LMS that was used to guide learners in utilizing cognitive, metacognitive, and motivational strategies of SRL in their learning. The log data from the self-regulated system, which was triangulated with self-reported data, assisted the students to practice SRL behaviors and enhanced their motivation and metacognition. In their study, Jayashanka et al. (2022) used the design science methodology to design a Moodle LMS plugin that helped students monitor their learning progress and academic performance. In the process, the technology-enhanced learning analytics dashboard was used to enhance students' engagement behaviors as a result of enhancing student motivation. Ustun et al. (2022) also conducted an experimental study that investigated the effect of learning analytics-based interventions on students SRL students and academic performance. The study, which was carried out on university students who were pursuing a computer science introductory course that lasted for 10 weeks, shows that feedback and recommendations rendered to students based on their inferred learning behaviors from log data help improve students' SRL skills and academic performance. In another related research, a deep learning network algorithm was used to develop a learning analytics model that was used to deliver real-time feedback in a computer-supported collaborative learning environment. The results reveal that offering real-time feedback improves collaborative interactions amongst students (Zheng et al., 2021). Our study, therefore, agrees with existing and current studies that offering SRL interventions in form of real-time feedback, prompts, and dashboard prescriptive analytics based on students' log data improves SRL skills, learner engagement, and interactions amongst online learners.

In conclusion, this study provides an innovative approach that uses EDM algorithms to develop an EDM model (plugin) integrated into LMS to enable learners actively get involved and engaged in their online learning activities thereby reinforcing their SRL skills. That constructivism and cognitive learning can be enhanced in LMS through the use of EDM algorithms that provide a mechanism for accessing and analyzing students' log data and proposing SRL interventions in real-time to the learners.

Table 5.1 presents the summary of the conclusions made from this study based on the study findings according to the research objectives.

Table 5.1: Summary of the Conclusions of the study based on Research Findings

Objective	Finding	Conclusion
To investigate the methods currently being used to measure and promote SRL in online learning environments	Self-report tools being used to measure SRL in both traditional classroom setup and online learning environments	Using self-report tools, students tend to overestimate their SRL levels and therefore unreliable
To investigate the current status of utilization of LMS features in promoting SRL in online learning.	LMS features are underutilized by learners and therefore fail to enable learners take control of their learning	Scaffold learners through the provision of SRL interventions enables students to plan, monitor and reflect on the learning practices
To develop and integrate into Moodle LMS an intelligent model that applies EDM to promote SRL strategies on LMSs	Agglomerative hierarchical clustering algorithm optimal in identifying SRL profiles for learners	Profiling of learners enables provision of targeted interventions to students with similar characteristics
To evaluate the model for its effectiveness in promoting SRL strategies on LMS	There was a significant difference between SRL strategies utilization by students who received EDM-based interventions and instructor-led intervention	The EDM model was effective in promoting students' SRL skills (cognitive and behavioral) on Moodle LMS
To establish students' perceived usefulness of the EDM model in promoting SRL on LMS.	The intervention messages helped learners reflect and improve on their undesirable learning behaviors	To enhance growth of students' SRL skills, intervention messages should be integrated in LMS using EDM

5.3 Recommendations

The recommendations made in this study are drawn from the conclusions of the study according to the research objectives. Table 5.2 presents the recommendations.

Table 5.2: Recommendations of the Study based of Resaerch Findings

Objective	Finding	Recommendation
To investigate the methods currently being used to measure and promote SRL in online learning environments	Self-report tools being used to measure SRL in both traditional classroom setup and online learning environments	Adoption on online methods for measuring SRL skills in online learning environments – unobtrusive, does not affect engagement
To investigate the current status of utilization of LMS features in promoting SRL in online learning.	LMS features are underutilized by learners and therefore fail to enable learners take control of their learning	Provision of feedback enable learners to pause and reflect on their learning behaviors before proceeding
To develop and integrate into Moodle LMS an intelligent model that applies EDM to promote SRL strategies on LMSs	Agglomerative hierarchical clustering algorithm optimal in identifying SRL profiles for learners	Integration of the EDM plugin into LMS systems for institutions of higher learning
To evaluate the model for its effectiveness in promoting SRL strategies on LMS	There was a significant difference between SRL strategies utilization by students who received EDM-based interventions and instructor-led intervention	Adoption of EDM-based interventions for promoting SRL as they are allow a learner to receive feedback based on their actual behaviors in an online learning environment

5.4 Research Contributions

This research aimed to examine the effectiveness of using educational data mining techniques in promoting self-regulated learning on learning management systems. As a result, this research has contributed knowledge in the following aspects;

1. **Technical Knowledge:** The study has contributed to the body of knowledge by implementing the EDM model that guides the design and provision of SRL interventions on LMS. The end artifact is an EDM tool/plugin that can be integrated into Moodle LMS as an intelligent tool that provides SRL interventions to support learner engagement in online learning. The Integration of the EDM plugin into Moodle LMS provides a new approach to how data stored on Moodle can be utilized to enhance online students' learning behavior. Furthermore, the plugin also enhances individualized learning support in online learning. This research, therefore, benefits researchers and practitioners in the Moodle Software Development Community.

   Since the model was developed and implemented using EDM algorithms, an empirical evaluation of EDM algorithms that was carried out to identify and recommend the optimal algorithm that profiles learners into similar clusters also contributes to technical knowledge in machine learning.
2. **Theoretical Knowledge:** This research also contributes to and enhances knowledge in the current literature in EDM and the e-learning community. The literature that has been published may be used by EDM and e-learning researchers and practitioners to identify future research opportunities. Additionally, the research conceptual framework and the reusable research instruments designed and used in this research may also enhance research in the areas of EDM, LMS, and SRL.
3. **E-learning Community:** This research also contributes knowledge to the e-learning community by creating an EDM plugin that will enhance students' support through the provision of interventions that enrich students' learning behavior. By integrating into Moodle LMS the EDM tool, it is intended that students' online learning will be more effective and satisfactory due to the automated support offered by the tool compared to the limited instructional support currently offered by instructors. The research also addresses the current challenges experienced by online learners as far as online support is concerned.

The contribution of the EDM tool further benefits Institutions of Higher Learning (IHL) and other education stakeholders on how to improve online learning through the provision of interventions that enhance students' involvement in online learning processes. Furthermore, the research provides a novel approach to improving Self-Regulated Learning in learning management systems.

Finally, the analysis of the LMS data about Self-Regulated learning provides a novel approach to measuring and promoting students' self-regulatory strategies in online learning environments. The integration of the EDM plugin into Moodle LMS provides an innovative approach to supporting students in ODEL learning environments especially those with undesirable online learning behaviors and patterns and hence enhancing instructional support which could not have been effectively provided due to the limited number of instructors in higher institutions of learning.

5.5 Policy Implications

In this study, we investigated the effectiveness of using EDM-based interventions in promoting SRL for online learners. The findings provide an opportunity of enhancing students' online learning experience. The findings also reveal that intelligent pedagogical agents designed under learning theories such as constructivism and cognitive theories and embedded into LMS allow and place learners at the center of their learning hence actively getting involved in their online courses. The following are policy implications that can be implemented by the respective stakeholders;

a. For Institutions of Higher Education

Through the findings arising from this research, we recommend that Institutions of Higher Education adopt the use of the EDM plugin developed and implemented in this study. Adopting and integrating the EDM plugin into LMSs currently being utilized by universities and colleges will enhance the delivery of online courses. Since the plugin will be proposing interventions to learners based on their actual learning behavior, there will be enhanced student support for online courses. Hence, increased learner motivation, satisfaction, and completion rates.

b. For Commission for University Education

The Commission for University Education is mandated in overseeing and ensuring quality teaching in Institutions of Higher Education in Kenya. One of the goals of the Commission for

University Education is to set up an open university in Kenya. It is therefore recommended that based on the findings from this study, as the Commission sets to plan and implement this goal, the adoption and use of the EDM plugin will enable the commission to ensure the following objectives are met.

i. Delivery of interactive online programs which meet the instructional needs of the learners
ii. The use of the EDM plugin in the LMS will provide automated interventions to students hence overcoming the limited instructor support in online courses.

c. For the Ministry of Education, Science, and Technology (MoEST)

The findings arising from this research will guide the Ministry of Education, Science, and Technology (MoEST) in implementing the Competency-Based Curriculum (CBC) by creatively implementing education reforms that will ensure learning/teaching is learner-centered. The adoption of enhanced interactive online learning environments that incorporates the EDM plugin will ensure that learning and support to the learners are individualized. MoEST should therefore envision the greater outcome of ensuring that education institutions in Kenya build and implement intelligent e-learning platforms at all levels of learning in the wake of the CBC era.

5.6 Suggestions for Future Research

The following areas have been suggested for further research to extend the findings in this study.

First, the plugin can be evaluated by varying contextual conditions to examine the effect of SRL interventions on students' SRL skills and academic performance. For example, the EDM plugin can be used to provide SRL interventions using multiple courses from other fields such as social sciences. Moreover, future studies can consider using courses that contribute to students' end-of-semester grades. These will provide different contexts to observe the effect of the interventions especially when students know that they have to study the course up to the end.

Secondly, since contextual limitations may hinder the generalization of the findings in this research since the study was carried out in selected institutions of higher learning from a developing country that could be facing various challenges such as inadequate access to internet access, further research can be carried out on students from developing countries where students do not face problems on internet access, there is a need to conduct an evaluation of the EDM plugin in other institutions of higher learning outside Kenya, especially in developed countries where challenges that online students in developing countries face are divergent.

Thirdly, this research was offered during the covid-19 pandemic that led to an unpredictable period where students stayed off-campus for a long time. As a result, the course that was offered in this study was not a regular university core course. This may have had on the students' motivation and commitment and seriousness with which they undertook their engagement, especially toward the end of the course when it got to too difficult sections such as practical sessions on machine learning, model development, and deployment. Future research should consider researching a course that is regulated by a university curriculum. Additionally, since this was an ICT-related course, the participants may have had prior knowledge on the use of computer systems such as LMS. It would therefore be interesting to see how such a study pans out with non-ICT courses such as social sciences and arts courses.

Lastly, this study took place for 12 weeks which may be a shorter period to observe the development of SRL skills and their long-term effect on academic performance. Future studies should examine the long-term impact of SRL interventions on students' growth of SRL strategies as well as on academic performance by conducting longitudinal studies where the levels of SRL will be inferred from level to level over a long period of learning hence affording the limitation of using self-report tools.

REFERENCES

Ainscough, L., Leung, R., Colthorpe, K., & Langfield, T. (2019). Characterizing university students' self-regulated learning behavior using dispositional learning analytics. *5th International Conference on Higher Education Advances (HEAd'19)*, 233–241. https://doi.org/10.4995/head19.2019.9153

Alario-Hoyos, C., Estévez-Ayres, I., Pérez-Sanagustín, M., Kloos, C. D., & Fernández-Panadero, C. (2017). Understanding learners' motivation and learning strategies in MOOCs. *International Review of Research in Open and Distance Learning*, *18*(3), 119–137. https://doi.org/10.19173/irrodl.v18i3.2996

Alasalmi, T. (2021). Students Expectations on Learning Analytics: Learning Platform Features Supporting Self-regulated Learning. *Proceedings Ofthe 13th International Conference on Computer Supported Education (CSEDU 2021)*, *2*, 131–140. https://doi.org/10.5220/0010537101310140

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*(January), 13–49. https://doi.org/10.1016/j.tele.2019.01.007

Aljumily, R. (2017). Agglomerative Hierarchical Clustering: An Introduction to Essentials. (1) Proximity Coefficients and Creation of a Vector-Distance Matrix and (2) Construction of the Hierarchical Tree and a Selection of Methods. *Global Journal of Human-Social Science: G Linguististics & Education*, *16*(3).

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. Berry, A. Mohamed, & B. Yap (Eds.), *Supervised and Unsupervised lesrning for Data Science* (Issue January, pp. 3–21). Springer. https://doi.org/10.1016/b978-0-12-818946-7.00003-2

Amineh, R., & Asl, H. (2015). Review of Constructivism and Social Constructivism. *Journal of Social Sciences, Literature and Languages*, *1*(1), 9–16.

Araiba, S. (2020). Current Diversification of Behaviorism. *Perspectives on Behavior Science*, *43*(1), 157–175. https://doi.org/10.1007/s40614-019-00207-0

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2019). A Conceptual Model for Measuring and Supporting Self-Regulated Learning using Educational Data Mining on Learning

Management Systems. In P. Cunningham & M. Cunningham (Eds.), *IST-Africa 2019* (pp. 1–11). IEEE. https://doi.org/10.23919/ISTAFRICA.2019.8764852

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—Systematic review (2008–2018). *Research and Practice in Technology Enhanced Learning*, *15*(1), 6. https://doi.org/10.1186/s41039-020-00129-5

Araka, E., Maina, E., Gitonga, R., Oboko, R., & Kihoro, J. (2021). University Students' Perception on the Usefulness of Learning Management System Features in Promoting Self-Regulated Learning in Online Learning. *International Journal of Education and Development Using Information and Communication Technology (IJEDICT)*, *17*(1), 45–64.

Araka, E., Oboko, R., Maina, E., & Gitonga, R. (2022). Using Educational Data Mining Techniques to Identify Profiles in Self-Regulated Learning: An Empirical Evaluation. *International Review of Research in Open and Distributed Learning*, *23*(1), 131–162. https://doi.org/10.19173/irrodl.v22i4.5401

Arkorful, V. (2014). The role of e-learning, the advantages and disadvantages of its adoption in Higher Education. *International Journal of Education and Research*, *2*(12), 397–410.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals in Pudu: Using learning analytics to Increase Student Success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 267–270. https://doi.org/10.1145/2330601.2330666

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, *4*(1), 87–95. https://doi.org/10.1007/s11409-009-9035-7

Azevedo, R., & Witherspoon, A. M. (2009). Self-Regulated Learning with Hypermedia Handbook of Metacognition in Education. In *Handbook of Metacognition in Education* (pp. 319–339). https://doi.org/10.4324/9780203876428.ch7

Back, D. A., Behringer, F., Haberstroh, N., Ehlers, J. P., Sostmann, K., & Peters, H. (2016). Learning management system and e-learning tools: An experience of medical students' usage and expectations. *International Journal of Medical Education*, *7*, 267–273. https://doi.org/10.5116/ijme.57a5.f0f5

Bada, O. S. (2015). Constructivism Learning Theory: A Paradigm for Teaching and Learning. *IOSR Journal of Research & Method in Education (IOSR-JRME)*, *5*(6), 66–70. https://doi.org/10.9790/7388-05616670

Baker, R. S. (2016). Stupid Tutoring Systems, Intelligent Humans. *International Journal of Artificial Intelligence in Education*, *26*(2), 600–614. https://doi.org/10.1007/s40593-016-0105-0

Baker, R., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, *1*(1), 3–16. http://doi.ieeecomputersociety.org/10.1109/ASE.2003.1240314

Bakhshinategh, B., Zaiane, O., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, *23*, 537–553. https://doi.org/10.1007/s10639-017-9616-z

Baracaldo, N., Chen, B., Ludwig, H., Safavi, A., & Zhang, R. (2018). Detecting poisoning attacks on machine learning in IoT environments. *Proceedings - 2018 IEEE International Congress on Internet of Things, ICIOT 2018 - Part of the 2018 IEEE World Congress on Services*, 57–64. https://doi.org/10.1109/ICIOT.2018.00015

Barnard, L., Lan, W., & Paton, V. (2010). Profiles in Self-Regulated Learning in the Online Learning Environment. *International Review of Research in Open and Distance Learning*, *11*(1), 55–78. https://doi.org/10.3109/10826087809039264

Boekaerts, M. (2017). Emotions, Emotion Regulation, and Self-Regulation of Learning. *Handbook of Self-Regulation of Learning and Performance, 13109*. https://doi.org/10.4324/9780203839010.ch26

Boekaerts, M., & Cascallar, E. (2006). How Far Have We Moved Toward the Integration of Theory and Practice in Self-Regulation? *Educ Psychol Rev*, *18*, 199–210. https://doi.org/10.1007/s10648-006-9013-4

Bogarín, A., Cerezo, R., & Romero, C. (2018a). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1). https://doi.org/10.1002/widm.1230

Bogarín, A., Cerezo, R., & Romero, C. (2018b). Discovering learning processes using inductive miner: A case study with learning management systems (LMSs). *Psicothema 2018, 30*(3), 322–329. https://doi.org/10.7334/psicothema2018.116

Bosch, N., Crues, W., Henricks, G., Perry, M., Angrave, L., Shaik, N., Bhat, S., & Anderson, C. (2018). Modeling key differences in underrepresented students' interactions with an online STEM course. *TechMindSociety '18*, 1–6. https://doi.org/10.1145/3183654.3183681

Bouchet, F., Harley, J. M., Trevors, G. J., & Azevedo, R. (2013a). Clustering and Profiling Students According to their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning. *JEDM - Journal of Educational Data Mining*, *5*(1), 104–146.

Bouchet, F., Harley, J. M., Trevors, G. J., & Azevedo, R. (2013b). Clustering and Profiling Students According to their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning. *JEDM - Journal of Educational Data Mining*, *5*(1), 104–146.

Broadbent, J., & Fuller-Tyszkiewicz, M. (2018). Profiles in self-regulated learning and their correlates for online and blended learning students. *Educational Technology Research and Development*, *66*(6), 1435–1455. https://doi.org/10.1007/s11423-018-9595-9

Broadbent, J., Panadero, E., Lodge, J., & Barba, P. de. (2020). Technologies to enhance self-regulated learning in online and computer mediated learning environments. In M. Bishop, J. Elen, E. B. And, & V. Svihla (Eds.), *Handbook of Research in Educational Communications and Technology*. Springer.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). ClValid: An R package for cluster validation. *Journal of Statistical Software*, *25*(4), 1–22. https://doi.org/10.18637/jss.v025.i04

Cavalcanti, A., Dourado, R., Rodrigues, R., Alves, N., Silva, J., & Ramos, J. L. C. (2018). An Analysis of Self-Regulated Learning Behavioral Diversity in Different Scenarios in Distance Learning Courses. *Anais Do XXIX Simpósio Brasileiro de Informática Na Educação (SBIE 2018)*, *1*(Cbie), 1493–1502. https://doi.org/10.5753/cbie.sbie.2018.1493

Çebi, A., & Güyer, T. (2020). Students' interaction patterns in different online learning activities and their relationship with motivation, self-regulated learning strategy and learning performance. *Education and Information Technologies*. https://doi.org/10.1007/s10639-020-10151-1

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). Process mining for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education*, *32*(1), 74–88. https://doi.org/10.1007/s12528-019-09225-y

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers and Education*, *96*, 42–54. https://doi.org/10.1016/j.compedu.2016.02.006

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). Nbclust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1–36. https://doi.org/10.18637/jss.v061.i06

Chen, T., He, T., & Benesty, M. (2018). XGBoost: EXtreme Gradient Boosting. *R Package Version 0.71-2*, 1–4.

Chou, C.-Y., & Zou, N.-B. (2020). An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models. *International Journal of Educational Technology in Higher Education*, *17*(55). https://doi.org/10.1186/s41239-020-00233-y

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). Finding Traces of Self-Regulated Learning in Activity Streams. *International Conference on Learning Analytics and Knowledge*, 10. https://doi.org/10.1145/3170358.3170381

Coman, C., Țîru, L. G., Meseșan-Schmitz, L., Stanciu, C., & Bularca, M. C. (2020). Online teaching and learning in higher education during the coronavirus pandemic: Students' perspective. *Sustainability (Switzerland)*, *12*(24), 1–22. https://doi.org/10.3390/su122410367

Contreras, P., & Murtagh, F. (2015). Hierarchical clustering. *Handbook of Cluster Analysis*, 103–124. https://doi.org/10.1201/b19706

Creswell, J. (2003). Research Design, Second Edition. In D. Laughton, V. Novak, & D. Axelsen (Eds.), *Research Design, 2nd Edition*. Inter.

Crossley, S., Mcnamara, D., Paquette, L., Baker, R., & Dascalu, M. (2016). Combining Click-Stream Data with NLP Tools to Better Understand MOOC Completion. *The 6th*

International Conference on Learning Analytics & Knowledge (LAK), 6–14. https://doi.org/10.1145/2883851.2883931

Dabbagh, N., & Kitsantas, A. (2005). Using web-based pedagogical tools as scaffolds for self-regulated learning. *Instructional Science*, *33*(5–6), 513–540. https://doi.org/10.1007/s11251-005-1278-3

Dabbagh, N., & Kitsantas, A. (2013). Using Learning Management Systems as Metacognitive Tools to Support Self-Regulation in Higher Education Contexts. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies, Springer International Handbooks of Education* (pp. 197–211). https://doi.org/10.1007/978-1-4419-5546-3_47

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Devasia, M. T., Vinushree, M., & Hegde, M. V. (n.d.). *Prediction of Students Performance using Educational Data Mining*.

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: Systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, *28*(6), 557–573. https://doi.org/10.1111/j.1365-2729.2011.00476.x

Di Mitri, D., Scheffel, M., Drachsler, H., Borner, D., Ternier, S., & Specht, M. (2016). Learning pulse: Using wearable biosensors and learning analytics to investigate and predict learning success in self-regulated learning. *CEUR Workshop Proceedings*, *1601*, 34–39.

Doleck, T., Basnet, R. B., Poitras, E. G., & Lajoie, S. P. (2015). Mining learner-system interaction data: Implications for modeling learner behaviors and improving overlay models. *Journal of Computers in Education*, *2*(4), 421–447. https://doi.org/10.1007/s40692-015-0040-3

Dolores, T. (2007). Purposive Sampling as a Tool for Informant Selection. *Ethnobotany Research & Applications*, *5*, 147–158.

Duncan, T. G., Mckeachie, W. J., & Duncan, T. G. (2005). The Making of the Motivated Strategies for Learning Questionnaire The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*, *5*(c), 15991–16005. https://doi.org/10.1109/ACCESS.2017.2654247

Elsayed, A., Caeiro-Rodríguez, M., Mikic-Fonte, F., & Llamas-Nistal, M. (2019). Research in Learning Analytics and Educational Data Mining to Measure Self-Regulated Learning: A Systematic Review. *World Conference on Mobile and Contextual Learning*, 46–53. https://www.learntechlib.org/p/210600/

Elsayyad, A., Nassef, A., & Baareh, A. (2015). Cardiac arrhythmia classification using boosted decision trees. *International Review on Computers and Software*, *10*(3), 280–289. https://doi.org/10.15866/irecos.v10i3.5359

Engelmann, K., Bannert, M., & Melzner, N. (2021). Do self-created metacognitive prompts promote short- and long-term effects in computer-based learning environments ? *Research and Practice in Technology Enhanced Learning*, *16*(3).

Ganesh, H., & Christy, J. (2015). Applications of Educational Data Mining: A survey. *2015 IEEE International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS 2015)*. https://doi.org/10.1109/ICIIECS.2015.7192945

Garcia, R., Falkner, K., & Vivian, R. (2018). Systematic literature review: Self-Regulated Learning strategies using e-learning tools for Computer Science. *Computers and Education*. https://doi.org/10.1016/j.compedu.2018.05.006

Gašević, D., Jovanović, J., Pardo, A., & Dawson, S. (2017). Detecting Learning Strategies with Analytics: Links with Self-Reported Measures and Academic Performance. *Journal of Learning Analytics*, *4*(2), 113–128. https://doi.org/10.18608/jla.2017.42.10

Grider, C. (1993). Foundations of Cognitive Theory: A Concise Review. *Information Analysis*, 15.

Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). *Predicting Students Performance in Educational Data Mining*. https://doi.org/10.1109/ISET.2015.33

Hadullo, K., Oboko, R., & Omwenga, E. (2018). Status of e-learning quality in Kenya: Case of Jomo Kenyatta University of agriculture and technology postgraduate students. *International Review of Research in Open and Distance Learning*, *19*(1), 138–159. https://doi.org/10.19173/irrodl.v19i1.3322

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, *17*(2/3), 107–145. https://doi.org/10.1023/A:1012801612483

Halverson, L. R., & Graham, C. R. (2019). Learner engagement in blended learning environments: A conceptual framework. *Online Learning Journal*, *23*(2), 145–178. https://doi.org/10.24059/olj.v23i2.1481

Hashemyolia, S., Asmuni, A., Daud, S. M., Fauzi, A., Ayub, M., & Shah, J. A. (2014). Factors Affecting Students' Self Regulated Learning Using Learning Management System. *Middle-East Journal of Scientific Research 19 (Innovation Challenges in Multidiciplinary Research & Practice)*, *19*, 119–124. https://doi.org/10.5829/idosi.mejsr.2014.19.icmrp.18

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A Systematic Review of Deep Learning Approaches to Educational Data Mining. *Complexity*, *2019*. https://doi.org/10.1155/2019/1306039

Ho, T. (1995). Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, *1*, 278–282. https://doi.org/10.1109/ICDAR.1995.598994

Huang, H.-M. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, *33*(1), 27. https://doi.org/Article

Hyeoun-Ae, P. (2013). An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, *43*(2), 154–164. https://doi.org/10.4040/jkan.2013.43.2.154

Iwamoto, D. H., Hargis, J., Bordner, R., & Chandler, inani. (2017). Self-Regulated Learning as a Critical Attribute for Successful Teaching and Learning. *International Journal for the Scholarship of Teaching and Learning*, *11*(2). https://doi.org/10.20429/ijsotl.2017.110207

Jansen, R., Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners' self-regulated learning in Massive Open Online Courses. *Computers & Education*, *146*, 103771. https://doi.org/10.1016/j.compedu.2019.103771

Janson, A., Söllner, M., & Leimeister, J. M. (2017). Individual Appropriation of Learning Management Systems—Antecedents and Consequences. *AIS Transactions on Human-Computer Interaction*, *9*(3), 173–201. https://doi.org/10.17705/1thci.00094

Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, *48*(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Jayashanka, R., Hettiarachchi, E., & Hewagamage, P. (2022). Technology Enhanced Learning in Higher Education. *The Electronic Journal of E-Learning*, *20*(2), 151–170.

Jha, N. I., Ghergulescu, I., & Moldovan, A. N. (2019). OULAD MOOC dropout and result prediction using ensemble, deep learning and regression techniques. *CSEDU 2019 - Proceedings of the 11th International Conference on Computer Supported Education*, *2*(Csedu), 154–164. https://doi.org/10.5220/0007767901540164

Jo, I.-H., Park, Y., Yoon, M., & Sung, H. (2016). Evaluation of Online Log Variables that Estimate Learners' Time Management in a Korean Online Learning Context. *International Review of Research in Open and Distributed Learning*, *17*(1), 195–213.

Job, M. A. (2018). Data Mining Techniques Applying on Educational Dataset to Evaluate Learner Performance Using Cluster Analysis. *European Journal of Engineering Research and Science*, *3*(11), 25–31. https://doi.org/10.24018/ejers.2018.3.11.966

Kallio, A., & Tuimala, J. (2013). Data Mining. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 525–528). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_599

Karaoglan Yilmaz, F. G., & Yilmaz, R. (2020). Learning analytics as a metacognitive tool to influence learner transactional distance and motivation in online learning environments. *Innovations in Education and Teaching International*, *00*(00), 1–11. https://doi.org/10.1080/14703297.2020.1794928

Karaoğlan Yılmaz, F. G., Olpak, D. Y. Z., & Yılmaz, R. (2018). The Effect of the Metacognitive Support via Pedagogical Agent on Self-Regulation Skills. *Journal of Educational Computing Research*, *56*(2), 159–180. https://doi.org/10.1177/0735633117707696

Karaoglan-Yilmaz, F.-G., & Yilmaz, R. (2021). Learning Analytics Intervention Improves Students' Engagement in Online Learning. *Technology, Knowledge and Learning*, *0123456789*. https://doi.org/10.1007/s10758-021-09547-w

Khampirat, B. (2021). Validation of Motivated Strategies for Learning Questionnaire:Comparison of Three Competing Models. *International Journal of Instruction*, *14*(2), 609–626. https://doi.org/10.29333/iji.2021.14234a

Khanna, L., Narayan Singh, S., & Alam, M. (2016). *Educational Data Mining and its Role in Determining Factors Affecting Students Academic Performance: A Systematic Review*.

Khiat, H., & Vogel, S. (2022). A self-regulated learning management system: Enhancing performance, motivation and reflection in learning. *Journal of University Teaching & Learning Practice, 19*(2), 43–59. https://doi.org/10.53761/1.19.2.4

Kim, D., Yoon, M., Jo, I. H., & Branch, R. M. (2018). Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. *Computers and Education*, *127*, 233–251. https://doi.org/10.1016/j.compedu.2018.08.023

Kim, H., Hong, A., & Song, H.-D. (2019). The roles of academic engagement and digital readiness in students' achievements in university e-learning environments. *International Journal of Educational Technology in Higher Education*, *16*(21). https://doi.org/10.1186/s41239-019-0152-3

Kinnebrew, J., Loretz, K., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *JEDM-Journal of Educational Data Mining*, *5*(1), 190–219.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers and Education*, *104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *ACM International Conference Proceeding Series*, 170–179. https://doi.org/10.1145/2460296.2460330

Kumar, Ks., & Raju, G. (2018). Study on Different Cluster Validity Indices. *International Journal of Applied Engineering Research*, *13*(11), 9364–9376.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Data Descriptor: Open University Learning Analytics dataset. *Scientific Data*, *4*, 1–8. https://doi.org/10.1038/sdata.2017.171

Lee, D., Watson, S. L., & Watson, W. R. (2019). Systematic literature review on self-regulated learning in massive open online courses. In *Australasian Journal of Educational Technology* (p. 35).

Lee, J., Song, H.-D., & Jeong Hong, A. (2019). Exploring Factors, and Indicators for Measuring Students' Sustainable Engagement in e-Learning. *Sustainability 2019*, *11*, 985. https://doi.org/10.3390/su11040985

Lee, J.-E., & Recker, M. (2017). *Measuring Students' Use of Self-Regulated Learning Strategies from Learning Management System Data: An Evidence-Centered Design Approach*.

Lee, J.-K. (2008). The effects of self-regulated learning strategies and system satisfaction regarding learner ' s performance in e-learning environment. *Journal of Instructional Pedagogies*, *1*, 30–45.

Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in Human Behavior*, *32*, 313–323. https://doi.org/10.1016/j.chb.2013.07.051

Li, H., Flanagan, B., Konomi, ichi, & Ogata, H. (2018). Measuring Behaviors and Identifying Indicators of Self-Regulation in Computer-Assisted Language Learning Courses. *Research and Practice in Technology Enhanced Learning*, *13*, 19. https://doi.org/10.1186/s41039-018-0087-7

Li, K., Yu, N., Li, P., Song, S., Wu, Y., Li, Y., & Liu, M. (2017). Multi-label spacecraft electrical signal classification method based on DBN and random forest. *PLoS ONE*, *12*(5), 1–19. https://doi.org/10.1371/journal.pone.0176614

Liu, D. Y. T., Bartimote-Aufflick, K., Pardo, A., & Bridgeman, A. J. (2017). Data-driven personalization of student learning support in higher education. *Studies in Systems, Decision and Control*, *94*, 143–169. https://doi.org/10.1007/978-3-319-52977-6_5

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). *Understanding of Internal Clustering Validation Measures*. https://doi.org/10.1109/ICDM.2010.35

Lodge, J. M., Panadero, E., Broadbent, J., & De Barba, P. G. (2019). Supporting self-regulated learning with learning analytics. *Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers*, *October*, 45–55.

Louppe, G. (2014). *Understanding Random Forests: From Theory to Practice* (Issue July). http://arxiv.org/abs/1407.7502

Luna, J. M., Castro, C., & Romero, C. (2017). MDM tool: A data mining framework integrated into Moodle. *Computer Applications in Engineering Education*, *25*(1), 90–102. https://doi.org/10.1002/cae.21782

Madni, H. A., Anwar, Z., & Shah, M. A. (2017). Data mining techniques and applications—A decade review. *ICAC 2017 - 2017 23rd IEEE International Conference on Automation and Computing: Addressing Global Challenges through Automation and Computing*, *January 2018*. https://doi.org/10.23919/IConAC.2017.8082090

Maina, E. M., Oboko, R. O., & Waiganjo, P. W. (2017). Using Machine Learning Techniques to Support Group Formation in an Online Collaborative Learning Environment. *International Journal of Intelligent Systems and Applications*, *9*(3), 26–33. https://doi.org/10.5815/ijisa.2017.03.04

Makokha, G. L., & Mutisya, D. N. (2016). Status of E-learning in public Universities in Kenya. *International Review of Research in Open and Distance Learning*, *17*(3), 341–359. https://doi.org/10.19173/irrodl.v17i3.2235

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*, 179–196. https://doi.org/10.1016/j.chb.2017.11.011

Mandler, G. (2002). Origins of the cognitive (r)evolution. *Journal of the History of the Behavioral Sciences*, *38*(4), 339–353. https://doi.org/10.1002/jhbs.10066

Manzanares, M., Sánchez, M., García, O., & Díez-Pastor, J. (2017). How do B-learning and learning patterns influence learning outcomes? *Frontiers in Psychology*, *8*(May 2017), 1–13. https://doi.org/10.3389/fpsyg.2017.00745

Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. *ACM International Conference Proceeding Series*, 461–470. https://doi.org/10.1145/3303772.3303787

McBroom, J., Yacef, K., & Koprinska, I. (2020). DETECT: A hierarchical clustering algorithm for behavioural trends in temporal educational data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12163 LNAI*, 374–385. https://doi.org/10.1007/978-3-030-52237-7_30

Montgomery, A. P., Mousavi, A., Carbonaro, M., Hayward, D. V., Dunn, W., & Montgomery, A. (2019). Using learning analytics to explore self-regulated learning in flipped blended learning music teacher education. *British Journal of Educational Technology*, *50*(1), 114–127. https://doi.org/10.1111/bjet.12590

Mouton, J. P., Ferreira, M., & Helberg, A. S. J. (2020). A comparison of clustering algorithms for automatic modulation classification. *Expert Systems with Applications*, *151*, 113317. https://doi.org/10.1016/j.eswa.2020.113317

Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction*, *58*(2018), 1–11. https://doi.org/10.1016/j.learninstruc.2018.04.011

Muuro, E., Wagacha, W., Oboko, R., & Kihoro, J. (2014). Students'perceived challenges in an online collaborative learning environment: A case of higher learning institutions in Nairobi, Kenya. *International Review of Research in Open and Distance Learning*, *15*(6), 132–161. https://doi.org/10.19173/irrodl.v15i6.1768

Muuro, M. E., & Kihoro, J. M. (2017). Learner experience of e-learning mode in institutions of higher learning: A case of Kenyan Universities. *2017 IST-Africa Week Conference, IST-Africa 2017*, 1–9. https://doi.org/10.23919/ISTAFRICA.2017.8102397

Naeem, S., & Wumaier, A. (2018). Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K. *International Journal of Computer Applications*, *182*(31), 7–14. https://doi.org/10.5120/ijca2018918234

Nam, C., & Smith-Jackson, T. (2007). Web-Based Learning Environment: A Theory-Based Design Process for Development and Evaluation. *Journal of Information Technology Education*, *6*, 23–43. https://doi.org/10.28945/200

Núñez, C., Romera, E., Magno, C., & Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology*, *8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Nussbaumer, A., Hillemann, E.-C., & Albert, D. (2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, *2*(1), 101–133.

Nussbaumer, A., Hillemann, E.-C., Gütl, C., & Albert, D. (2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1), 72–100. https://doi.org/10.18608/jla.2015.21.5

Oboko, R. (2012). *Adaptive learning and Metacognitive Regulation Supportfor Ill-structured problem solving processes in web-based learning*.

Offermann, P., Levina, O., Schönherr, M., & Bub, U. (2009). Outline of a design science research process. *Proceedings of the 4th International Conference on Design Science Research in*

Information Systems and Technology - DESRIST '09, *January*, 1. https://doi.org/10.1145/1555619.1555629

Oloruntoba, S. A., & Akinode, J. L. (2017). STUDENT ACADEMIC PERFORMANCE PREDICTION USING SUPPORT VECTOR MACHINE. *International Journal of Engineering Sciences & Research Technology*, *588*. https://doi.org/10.5281/zenodo.1130905

Osman, A., Ahmed, N., Chow, M., Huang, Y., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

Otieno, J. (2016). *State to set up open and e-learning university: Kenya—The Standard*. Standard Newspaper. https://www.standardmedia.co.ke/article/2000207884/state-to-set-up-open-and-e-learning-university

Palomino, G., Silveira, R., & Nakayama, M. (2014). An intelligent LMS model based on intelligent tutoring systems. *International Conference on Intelligent Tutoring Systems, ITS2014*, 567–574. https://doi.org/10.1007/978-3-319-07221-0_72

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*. https://doi.org/10.1080/00313831.2015.1066436

Papamitsiou, Z., & Economides, A. (2014). International Forum of Educational Technology & Society Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence Linked references are available on JSTOR for this article: Learning Analytics and Ed. *Educational Technology & Society*, *17*(4), 49–64.

Pardo, A., Bartimote-Aufflick, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martinez-Maldonaldo, R., Mirriahi, N., Moskal, A. C. M., Schulte, J., Siemens, G., & Vigentini, L. (2018). OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Journal of Learning Analytics*, *5*(3), 235–249. https://doi.org/10.18608/jla.2018.53.15

Park, J., Yu, R., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). Understanding Student Procrastination via Mixture Models. *The 11th International Conference on Educational Data Mining (EDM)*, 187–197.

Peach, R. L., Yaliraki, S. N., Lefevre, D., & Barahona, M. (2019). Data-driven unsupervised clustering of online learner behaviour. *Npj Science of Learning*, *4*(1), 1–11. https://doi.org/10.1038/s41539-019-0054-0

Peng, C.-Y., Lee, K., & Ingersoll, G. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, *96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Pérez-Sanagustín, M., Sapunar-Opazo, D., Pérez-Álvarez, R., Hilliger, I., Bey, A., Maldonado-Mahauad, J., & Baier, J. (2020). A MOOC-based flipped experience: Scaffolding SRL strategies improves learners' time management and engagement. *Computer Applications in Engineering Education*, *January*, 1–19. https://doi.org/10.1002/cae.22337

Pintrich, P. (2000). The Role of Goal Orientation in Self-Regulated Learning. *Handbook of Self-Regulation*, 451–502.

Pintrich, P. (2004). A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1016/j.compag.2016.08.024

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*.

Pintrich, P., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. *The Development of Achievement Motivation*, *1446*(2), 249–284. https://doi.org/10.1016/b978-012750053-9/50003-6

Ratnapala, I., Ragel, R., Deegalla, S., Ratnapala, I. P., Ragel, R. G., & Deegalla, S. (n.d.). *Students Behavioural Analysis in an Online Learning Environment Using Data Mining Optical Font Recognition using Machine Learning methods for Sinhalese Text View project PhD research of S. Vidanagamachchi under my supervision View project Students Behavio*.

Rodriguez, A., Vázquez Barreiros, B., Lama, M., Gewerc, A., & Mucientes, M. (2014). Using a Learning Analytics Tool for Evaluation in Self-Regulated Learning. *2014 IEEE Frontiers in Education Conference*, 2484–2491.

Rodriguez, F., Rivas, M. J., Yu, R., Warschauer, M., Park, J., & Sato, B. K. (2019). Utilizing learning analytics to map students' self-reported study strategies to click behaviors in STEM courses. *The 9th International Conference on Learning Analytics & Knowledge (LAK)*, 456–460. https://doi.org/10.1145/3303772.3303841

Rokach, L., & Maimon, O. (2008). Decision Trees. *Data Mining and Knowledge Discovery Handbook*, 165–192. https://doi.org/10.1007/978-3-540-75859-4_5

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12. https://doi.org/10.18608/jla.2015.21.2

Romero, C., López, M. I., Luna, J. M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers and Education*, *68*, 458–472. https://doi.org/10.1016/j.compedu.2013.06.009

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(2007), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs Data Mining Knowl Discov*, *3*, 12–27. https://doi.org/10.1002/widm.1075

Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(1). https://doi.org/10.1002/widm.1187

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, *51*(1), 368–384. https://doi.org/10.1016/j.compedu.2007.05.016

Roth, A., Ogrin, S., & Schmitz, B. (2016a). Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability*. https://doi.org/10.1007/s11092-015-9229-2

Roth, A., Ogrin, S., & Schmitz, B. (2016b). Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability*, *28*(3), 225–250. https://doi.org/10.1007/s11092-015-9229-2

Rowe, F. A., & Rafferty, J. A. (2013). Instructional Design Interventions for Supporting Self-Regulated Learning: Enhancing Academic Outcomes in Postsecondary E-Learning

Environments. In *MERLOT Journal of Online Learning and Teaching* (Vol. 9, pp. 590–601).

Saa, A. A. (2016). Educational Data Mining & Students' Performance Prediction. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 7).

Şahin, M., & Yurdugül, H. (2020). Learners' Needs in Online Learning Environments and Third Generation Learning Management Systems (LMS 3.0). *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-020-09479-x

Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., Díez-Pastor, J. F., & García-Osorio, C. I. (2019). Does the use of learning management systems with hypermedia mean improved student learning outcomes? *Frontiers in Psychology*, *10*(FEB), 1–14. https://doi.org/10.3389/fpsyg.2019.00088

Saks, K., & Leijen, Ä. (2014). Distinguishing Self-Directed and Self-Regulated Learning and Measuring them in the E-learning Context. *International Conference on Education & Educational Psychology 2013 (ICEEPSY 2013)*, *112*, 190–198. https://doi.org/10.1016/j.sbspro.2014.01.1155

Sathya, R., & Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, *2*(2), 34–38.

Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist*, *45*(4), 258–266. https://doi.org/10.1080/00461520.2010.515936

Schubert, E. (2020). HACAM: Hierarchical Agglomerative Clustering around Medoids – and its limitations. *CEUR Workshop Proceedings*, *2993*, 191–204.

Schumacher, C., & Ifenthaler, D. (2020). Investigating prompts for supporting students' self-regulation – A remaining challenge for learning analytics approaches? *The Internet and Higher Education*, 100791. https://doi.org/10.1016/j.iheduc.2020.100791

Schunk, D. (2005). Self-Regulated Learning: The Educational Legacy of Paul R. Pintrich. *Educational Psychologist*, *40*(2), 85–94. https://doi.org/10.1207/s15326985ep4002_3

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. (2018). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 105512. https://doi.org/10.1016/j.chb.2018.05.004

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, *18*(54). https://doi.org/10.1186/s41239-021-00292-9

Siadaty, M. (2016). Trace-Based Microanalytic Measurement of Self-Regulated Learning Processes Moray House School of Education and School of Informatics. *Journal of Learning Analytics*, *3*(1), 183–214.

Sin, K., & Muthu, L. (2015). APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS-A LITERATURE REVIEW. *ICTACT JOURNAL ON SOFT COMPUTING: SPECIAL ISSUE ON SOFT COMPUTING MODELS FOR BIG DATA*, 4. https://doi.org/10.21917/ijsc.2015.0145

Skinner, B. F. (1938). The behavior of organisms: An experimental analysis. In *The behavior of organisms: An experimental analysis.* Appleton-Century.

Sukhija, K., Jindal, M., & Aggarwal, N. (2016). The recent state of educational data mining: A survey and future visions. *Proceedings of the 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education, MITE 2015*, 354–359. https://doi.org/10.1109/MITE.2015.7375344

Sun, Z., Lu, L., & Xie, K. (2016). The effects of self-regulated learning on students' performance trajectory in the flipped math classroom. *Proceedings of International Conference of the Learning Sciences (ICLS 2016)*, *1*, 66–73.

Susheelamma, K. H., & Ravikumar, K. M. (2019). Student risk identification learning model using machine learning approach. *International Journal of Electrical and Computer Engineering*, *9*(5), 3872–3879. https://doi.org/10.11591/ijece.v9i5.pp3872-3879

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, *19*(12). https://doi.org/10.1186/s41239-021-00313-7

Sweller, J. (2011). Cognitive Load Theory. *Psychology of Learning and Motivation - Advances in Research and Theory*, *55*, 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Syuhada, R., Mawengkang, H., & Lydia, M. S. (2020). Analysis of performances k-nearest neighbor for regulate learning. *IOP Conference Series: Materials Science and Engineering*, *725*(3). https://doi.org/10.1088/1757-899X/725/1/012132

Tam, M. (2000). Constructivism, Intructional Design, and Technology: Implications for Transforming Distance Learning. *Educational Technology & Society*, *3*(2), 50–60.

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology*, *46*(3), 472–487. https://doi.org/10.1111/bjet.12274

Trevors, G., Feyzi-Behnagh, R., Azevedo, R., & Bouchet, F. (2016). Self-regulated learning processes vary as a function of epistemic beliefs and contexts: Mixed method evidence from eye tracking and concurrent and retrospective reports. *Learning and Instruction*, *42*, 31–46. https://doi.org/10.1016/j.learninstruc.2015.11.003

Triquet, K., Peeters, J., & Lombaerts, K. (2017). *Self-Regulated Learning Online: Benefits, Empirical Foundations, Multi-level, and Multi-modal Promotion & Evaluation thereof for Teacher Professional Development. Contributing SRL Part to Teach-UP. A policy experimentation co-funded by Erasmus+. Deliverab*. Department of Educational Sciences, Vrije Universiteit Brussel.

Ustun, A., Zhang, K., Karaoğlan-Yilmaz, F., & Yilmaz, R. (2022). Learning analytics based feedback and recommendations in flipped classrooms: An experimental study in higher education. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2022.2040401

Valarmathy, N., & Krishnaveni, S. (2019). Performance evaluation and comparison of clustering algorithms used in educational data mining. *International Journal of Recent Technology and Engineering*, *7*(6), 103–112.

Valdiviezo, P., Reátegui, R., & Sarango, M. (2013). Student Behavior Patterns in a Virtual Learning Environment. *Eleventh LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2013)*, 1–8.

Valle, A., Núñez, C., Cabanach, R., González-pienda, J., Rodríguez, S., Rosário, P., Cerezo, R., & Muñoz-cadavid, M. (2008). Self-regulated profiles and academic achievement. *Psicothema*, *20*, 724–731.

van Alten, D., Phielix, C., Janssen, J., & Kester, L. (2020). Effects of self-regulated learning prompts in a flipped history classroom. *Computers in Human Behavior*, *108*. https://doi.org/10.1016/j.chb.2020.106318

van Merriënboer, J., & de Bruin, A. (2019). Cue-based facilitation of self-regulated learning: A discussion of multidisciplinary innovations and technologies. *Computers in Human Behavior*, *100*(July), 384–391. https://doi.org/10.1016/j.chb.2019.07.021

Viberg, O., Khalil, M., & Baars, M. (2020). Self-Regulated Learning and Learning Analytics in Online Learning Environments: A Review of Empirical Research. *The 10th International Conference on Learning Analytics & Knowledge (LAK'20)*, 11. https://doi.org/10.1145/3375462.3375483

von Glasersfeld, E. (2014). Piaget and the radical constructivist epistemology. *Costruttivismi*, *1*, 94–107. https://doi.org/10.23826/2014.02.094.107

Vovides, Y., Sanchez-Alonso, S., Mitropoulou, V., & Nickmans, G. (2007). The use of e-learning course management systems to support learning strategies and to improve self-regulated learning. *Educational Research Review*, *2*, 64–74. https://doi.org/10.1016/j.edurev.2007.02.004

Wandler, J. B., & Imbriale, W. J. (2017). Promoting Undergraduate Student Self-Regulation in Online Learning Environments. *Online Learning*, *21*(2), 1–16. https://doi.org/10.24059/olj.v21i2.881

Whipp, J. L., & Chiarelli, S. (2004). Self-Regulation in a Web-Based Course: A Case Study. *Educational Technology Research and Development*, *52*(4), 5–22.

Winne, P. (2011). A Cognitive and Metacognitive Analysis of Self-Regulated Learning. *Handbook of Self-Regulation of Learning and Performance, 13109*, 15–32. https://doi.org/10.4324/9780203839010.ch2

Winne, P., & Baker, R. (2013). The Potentials of Educational Data Mining for Researching Metacognition, Motivation and Self-Regulated Learning. *Journal of Educational Data Mining*, *5*(1), 1–8. https://doi.org/10.1037/1082-989X.2.2.131

Winne, P. H., & Perry, N. E. (2000). Measuring Self-Regulated Learning. In *Handbook of Self-Resulation 5*.

Wirth, J. (2009). Promoting Self-Regulated learning Throgh Prompts. *Zeitschrift Fur Padagogische Psychologie*, *23*(2), 91–94. https://doi.org/10.1024/1010-0652.23.2.91

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G.-J., & Paas, F. (2018). Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. *International Journal of Human–Computer Interaction*, 1–18. https://doi.org/10.1016/j.cadmos.2010.11.005

Wong, J., Baars, M., de Koning, B., & Paas, F. (2020). Examining the use of Prompts to Facilitate Self-Regulated Learning in Massive Open Online Courses. *Computers in Human Behavior*, 105398. https://doi.org/10.1016/j.chb.2020.106596

Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers and Education*, *140*(2019), 103595. https://doi.org/10.1016/j.compedu.2019.103595

Xu, D., Huang, W., Wang, H., & Heales, J. (2014). Enhancing e-learning effectiveness using an intelligent agent-supported personalized virtual learning environment: An empirical investigation. *Information and Management*, *51*(4), 430–440. https://doi.org/10.1016/j.im.2014.02.009

Yilmaz, R., Karaoğlan Yılmaz, F. G., & Cakmak, E. (2017). The impact of transactive memory system and interaction platform in collaborative knowledge construction on social presence and self-regulation. *Interactive Learning Environments*, *25*(8), 949–969. https://doi.org/10.1080/10494820.2016.1224905

Yoon, M., Hill, J., & Kim, D. (2021). Designing supports for promoting self-regulated learning in the flipped classroom. *Journal of Computing in Higher Education*, *33*(2), 398–418. https://doi.org/10.1007/s12528-021-09269-z

Yot-Domínguez, C., & Marcelo, C. (2017). University students' self-regulated learning using digital technologies. *International Journal of Educational Technology in Higher Education*, *14*(38), 1–18. https://doi.org/10.1186/s41239-017-0076-8

Yu, R., Jiang, D., & Warschauer, M. (2018). Representing and predicting student navigational pathways in online college courses. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale - L@S '18*, 1–4. https://doi.org/10.1145/3231644.3231702

Zheng, J., Xing, W., Zhu, G., Chen, G., Zhao, H., & Xie, C. (2020). Profiling self-regulation behaviors in STEM learning of engineering design. *Computers & Education*, *143*(2020), 103669. https://doi.org/10.1016/j.compedu.2019.103669

Zheng, L., Niu, J., & Zhong, L. (2021). Effects of a learning analytics-based real-time feedback approach on knowledge elaboration, knowledge convergence, interactive relationships and group performance in CSCL. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13156

Zhou, M., & Brown, D. (2015). Educational Learning Theories: 2nd Edition. In *Education Open Textbooks*. GALILEO Open learning Materials. https://oer.galileo.usg.edu/education-textbooks/1

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, *11*(4), 307–313. https://doi.org/10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501

APPENDICES

Appendix I: Publications

This appendix contains research work that has been published before the submission of the thesis.

Part of the research work in this thesis has been published in peer-reviewed journals before the submission of the thesis.

Araka, E., Oboko, R., Maina, E., & Gitonga, R. (2022). Using Educational Data Mining Techniques to Identify Profiles in Self-Regulated Learning: An Empirical Evaluation. *The International Review of Research in Open and Distributed Learning*, 23(1), 131-162. https://doi.org/10.19173/irrodl.v22i4.5401

Araka, E., Maina, E., Gitonga, R., Oboko, R., Kihoro, J. (2021). University Students' Perception on the Usefulness of Learning Management System Features in Promoting Self-Regulated Learning in Online Learning. *International Journal of Education and Development using ICT*, (17)1. http://ijedict.dec.uwi.edu/viewarticle.php?id=2850

Araka, E., Maina, E., Gitonga, R., Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments - systematic review (2008–2018). *Research and Practice in Technology Enhanced Learning*, 15(6), 1-21. https://link.springer.com/article/10.1186/s41039-020-00129-5

Part of the research work in this thesis has been presented at a conference before the submission of the thesis.

Araka, E., Maina, E., Gitonga, R., Oboko, R. (2019). A Conceptual Model for Measuring and Supporting Self-Regulated Learning using Educational Data Mining on Learning Management Systems. *IST-Africa Week Conference (IST-Africa) 2019*, Nairobi, Kenya, pp. 1-11, DOI: 10.23919/ISTAFRICA.2019.8764852. https://ieeexplore.ieee.org/document/8764852

Part of the research work in this thesis has been published in a book chapter before the submission of the thesis.

Araka, E., Oboko, R., Maina, E., & Gitonga, R. K. (2021). A Conceptual Educational Data Mining Model for Supporting Self-Regulated Learning in Online Learning Environments. In J. Keengwe, & Y. Tran (Ed.), *Handbook of Research on Equity in Computer Science in P-16 Education* (pp. 278-292). IGI Global. http://doi:10.4018/978-1-7998-4739-7.ch016

Appendix II: Research Permit to conduct research by NACOSTI

This appendix contains the research permit issued by National Commission for Science, Technology & Innovation (NACOSTI).

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **496829**

Date of Issue: **19/May/2021**

RESEARCH LICENSE

This is to Certify that Mr.. Eric Nyambiriga Araka of Kenyatta University, has been licensed to conduct research in Kajiado, Kiambu, Machakos, Nairobi on the topic: An Educational Data Mining Model for Promoting Self-Regulated Learning on Learning Management Systems for the period ending : 19/May/2022.

License No: **NACOSTI/P/21/10723**

496829

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix III: Approval Letter to conduct research at Kenyatta University

This appendix contains the approval letter to conduct research at Kenyatta University.

KENYATTA UNIVERSITY

OFFICE OF DEPUTY VICE-CHANCELLOR, RESEARCH, INNOVATION AND OUTREACH

Ref: KU/DVCR/RCR/VOL.3/312

P. O. Box 43844 – 00100
Nairobi. Kenya
Tel. 254-20-810901 Ext. 026
E-mail: dvc-rio@ku.ac.ke

Mr. Eric Araka,
Dept. of Computing & Info. Technology,
Kenyatta University,
NAIROBI

8th June, 2021

Dear Mr. Araka,

RE: **REQUEST TO COLLECT RESEARCH DATA AT KENYATTA UNIVERSITY**

This is with reference to your letter dated 15th February, 2021 requesting for authorization to collect research data at Kenyatta University on the topic **"An Educational Data Mining Model for Promoting Self Regulated Learning on Learning Management Systems"** *towards the PhD (Computer Science) degree of Kenyatta University.*

I am happy to inform you that the Vice-Chancellor has approved your request to collect data. It has been noted that your data will be collected from CIT students through an online survey.

Yours Sincerely,

Prof. F. Q. Gravenir
Deputy Vice-Chancellor
Research, Innovation & Outreach

cc. Vice-Chancellor
DVC, Administration

Appendix IV: Pre-study Questionnaire

This appendix contains the questionnaire used in the pre-study of this research

Students' Perception on Learning Management System Features in Promoting Self-Regulated Learning on online learning environments

Greetings,
My name is Eric Araka and I am pursuing my PhD in Computer Science from Kenyatta University.

I am conducting a research to investigate the extent to which LMS features and functionalities such as chats, forums, workshops, quizzes and wikis are being utilized in enhancing self-regulated learning skills. Self-Regulated Learning skills enable learners to take control in managing their own learning and assume an active role in achieving their academic goals.

We would like to request for about 10 minutes of your time to respond to the questions below. Your participation is important as it will help us identify ways of improving the way students are supported while studying online.

The information you will provide will be treated with utmost confidentiality and will only be used for academic purposes.

---

*Required

1. Gender *

Mark only one oval.

- Male
- Female

2. Age *

Mark only one oval.

- 15-25 Years
- 26-35 Years
- 36-45
- 46-55
- Above 56 Years

3. In which University are you currently studying? *

Mark only one oval.

- Kenyatta University
- Maseno University
- The Cooperative University of Kenya
- Mount Kenya University
- KCA University

4. What level of study are you currently in? *

Mark only one oval.

- Certificate
- Diploma
- Undergraduate
- Postgraduate

5. What programme are you enrolled in? (Indicate your answer on the space below for example Bachelor of Commerce, Bachelor of IT) *

____

6. What mode of learning are you enrolled in your study? *

Mark only one oval.

- Blended learning (online classes to supplement face-to-face classes)
- Fully Online (All classes are online)
- Part-time or Evening
- Full-Time student

7. How many hours per week do you spend studying online? *

Mark only one oval.

- 1-10 hours
- 11-20 hours
- 21-30 hours
- 31-40 hours
- 41 or more

8. What online communication tool(s) do instructors use most to communicate with students? *

Mark only one oval.

- Email
- Messages
- Chat
- Forum

9. What online communication tool(s) do you use most to communicate with fellow students? *

Mark only one oval.

- Email
- Messages
- Chat
- Forum

10. From your own experience, how can you rate the communication with your course instructors? *

Mark only one oval.

- Very Often
- Often
- Sometimes
- Rarely
- Never

11. Indicate how often you utilize the following features on the LMS during your online learning *

Mark only one oval per row.

	Very often	Often	Sometimes	Rarely	Never
Forums					
Chats					
Wikis					
Emails					
Messages					
Quizzes					
Blogs					
Workshops					

12. For what purpose do you use the following LMS features during an online course? *

Tick all that apply.

	Seek help from course instructors	Seek help from other students	Meet course goals	Meet strategy goals	I have never used the tool	I am not aware on the tool
Chat						
Forums						
Messages						

13. For what purpose do you use Wikis during an online course? *

Tick all that apply.

- [ ] Share knowledge with others students
- [ ] Show to the instructor the evidence of assignment progress
- [ ] Collaborate with other students
- [ ] I have never used wikis
- [ ] I am not aware what wikis are

14. For what purpose do you use Quizzes during an online course? *

Tick all that apply.

- [ ] Meet the assessment requirements
- [ ] Monitor your learning process
- [ ] Evaluate yourself
- [ ] I have never used Quizzes
- [ ] I am not aware what Quizzes are

15. For what purpose do you use workshops during an online course? *

Tick all that apply.

- [ ] Share content for your colleagues to assess you
- [ ] I have never used workshops
- [ ] I am not aware what workshops are

16. From your own experience, what features and tools do the instructors use to monitor your learning progress? *

Tick all that apply.

- [ ] Quizzes
- [ ] Assignments

Other: [ ] ____________________

17. In your own opinion, what hinders you from actively being involved in online learning? Check on any of the following that apply *

Tick all that apply.

- [ ] Lack of instructor guidance
- [ ] Lack of individualized feedback on your learning habits
- [ ] Lack of peer interaction
- [ ] Lack of interaction with course instructors
- [ ] Lack of adequate internet
- [ ] Lack of adequate learning

Other: [ ] ____________________

18. How often do you receive feedback that supports you to improve learning habits in terms of utilizing LMS features to engage you in learning? *

Mark only one oval.

- ( ) Very often
- ( ) Often
- ( ) Sometimes
- ( ) Rarely
- ( ) Never

19. From whom does the feedback come from? Select from the options below *

Mark only one oval.

- ( ) System
- ( ) Instructor
- ( ) Both System and Instructor
- ( ) Not aware of the source of feedback
- ( ) Never receive any feedback

20. From your own experience, what LMS feature(s) do instructors use most to monitor your learning progress? *

Tick all that apply.

- [ ] Emails
- [ ] Messages
- [ ] Badges
- [ ] Q&A Forums
- [ ] Progress Bar

Other: [ ] ______________________

21. Among the following instructor roles, which areas do you feel there is little support from course instructors? *

Tick all that apply.

- [ ] Individualized feedback
- [ ] Real time feedback
- [ ] Guided learning
- [ ] Study hints
- [ ] Prompts guiding you on your study habits
- [ ] Instructional help
- [ ] Provision of learning materials

Other: [ ] ______________________

22. From the following list of Self-Regulated Learning strategies select the ones you have utilized most in an online course *

Tick all that apply.

- [ ] Time management by using LMS features to plan time use e.g. frequent logins and when to respond to various activities shared by instructors
- [ ] Self-evaluation through the use of quizzes and self-assessment tools on LMS
- [ ] Self-monitoring through the use of tracking or journaling tools to monitor learning progress
- [ ] Help Seeking by seeking assistance from peers or instructors through Q&A, chats and forums
- [ ] Organization by scheduling learning activities for the day or week
- [ ] Effort management by actively participating in online learning by having frequent logins and reminders to view learning materials. Know the effort to put towards achieving you course goals
- [ ] None of the above
- [ ] I am not aware of the SRL strategies

Other: [ ] ______________________

23. How often do you get feedback on performance in submitted of assignments and quizzes *

Mark only one oval.

- Very often
- Often
- Sometimes
- Rarely
- Never

24. From what you have experienced during an online course study indicate your opinion on the following issues *

Mark only one oval per row.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I am satisfied with the level of support I receive from course instructors					
The instructional feedback from instructors takes too long					
I am satisfied with the level of individualized guidance and support I receive from instructors					
I am satisfied that my individual needs are catered for by studying online					
I am satisfied with the instructional support I receive from course instructors					
I am satisfied with the regular guidelines on how to utilize online features during the learning process					

25. Among the following tasks and tools tick the one(s) you ever utilized to manage your learning *

Tick all that apply.

	Forums	Chats	Wikis	Emails	Workshops	Quizzes	None of the tools
Manage your time while studying online	☐	☐	☐	☐	☐	☐	☐
Monitor your study strategies	☐	☐	☐	☐	☐	☐	☐
Evaluate yourself in terms of progress of study to achieve the course goals	☐	☐	☐	☐	☐	☐	☐
Searching for help from peers and course instructors	☐	☐	☐	☐	☐	☐	☐
Effort management in terms of knowing where to commit more time and effort	☐	☐	☐	☐	☐	☐	☐
To plan an organise your learning activities	☐	☐	☐	☐	☐	☐	☐

26. Among the following SRL strategies tick the approach that will be most appropriate in supporting you to manage your learning process *

Tick all that apply.

	Use of Prompts	Use of Feedback	Automated system	Progress bar	Use LMS calender	To do list	Instructor Support
Time Management	☐	☐	☐	☐	☐	☐	☐
Self-Monitoring	☐	☐	☐	☐	☐	☐	☐
Self-Evaluation	☐	☐	☐	☐	☐	☐	☐
Help Seeking	☐	☐	☐	☐	☐	☐	☐
Organisation	☐	☐	☐	☐	☐	☐	☐
Effort Management	☐	☐	☐	☐	☐	☐	☐

27. What additional features do you think the e-learning system should have for supporting your individual learning

28. Briefly describe your experience in online courses in terms of the support you receive from instructors

Appendix V: Pretest MSLQ Questionnaire

This appendix contains the post-test that was used to measure students' self-regulated learning skills after the experiment.

Data Science with Python: Pre-enrollment Survey

First, thank you for registering to train on the Data Science with Python Course. Before proceeding to enroll you in the course we would like to ask questions about the way you learn as it will help us identify ways of providing support for online courses such as Data Science with Python.

We therefore request that you respond to the question below and be honest and accurate as possible. Please rate the questions based on the way you learn such courses as Data Science with Python.

The rating is based on a 7-point scale where 1 = not at all true of me to 7 = very true of me.

*Required

1. Email *

2. If I study in appropriate ways, then I will be able to learn the material in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

3. It is my own fault if I don't learn the material in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

4. If I try hard enough, then I will understand the course material. *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

5. If I don't understand the course material, it is because I didn't try hard enough. *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

6. I believe I will receive an excellent grade in this class. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

7. I'm certain I can understand the most difficult material presented in the readings for this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

8. I'm confident I can understand the basic concepts to be taught in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

9. I'm confident that I can understand the most complex material presented by the instructor in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

10. I'm confident that I can do an excellent job on the assignments and tests in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

11. I expect to do well in this class. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

12. I'm certain that I can master the skills that will be taught in this class. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

13. Considering the difficulty of this course, the instructor, and my skills, I think I will do well in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

14. When I study for this course, I practice saying the material to myself over and over. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

15. When studying for this course, I read my class notes and the course materials over and over again. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

16. I memorize key words to remind me of important concepts in this class. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

17. I make lists of important terms for this course and memorize the lists. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

18. I usually study in a place where I can concentrate on my coursework. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

19. I make good use of my study time for this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

20. I find it hard to stick to a study schedule. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

21. I have a regular place set aside for studying. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

22. I make sure I keep up with the weekly readings and assignments for this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me								Very true of me

23. I attend class regularly. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me								Very true of me

24. I often find that I don't spend very much time on this course because of other activities. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me								Very true of me

25. I rarely find time to review my notes or readings before an exam. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me								Very true of me

26. I often feel so lazy or bored when I study for this class that I quit before I finish what I planned to do. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me								Very true of me

27. I work hard to do well in this class even if I don't like what we are doing. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

28. When coursework is difficult, I give up or only study the easy parts. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

29. Even when course materials are dull and uninteresting, I manage to keep working until I finish. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

30. Even if I have trouble learning the material in this course, I try to do the work on my own, without help from anyone. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

31. I ask the instructor to clarify concepts I don't understand well. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

32. When I can't understand the material in this course, I ask another student in this class for help. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

33. I try to identify students in this class whom I can ask for help if necessary. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

Appendix VI: Posttest MSLQ Questionnaire

This appendix contains the post-test that was used to measure students' self-regulated learning skills after the experiment.

Data Science with Python: Post study Survey

We would like to thank you for enrolling and completing the Data Science with Python Course. In view of this, we would like to ask questions about the way you studied this course as it will help us identify ways of improving support to online courses.

We therefore request that you respond to the questions that follow and be as honest and accurate as possible. Please rate the questions based on the way you studied the Data Science with Python Course.

The rating is based on a 7-point scale where 1 = not at all true of me to 7 = very true of me.

*Required

1. Email *

2. I studied this course in appropriate ways and therefore I was able to learn the material that was provided *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

3. It was my own fault if I did not learn the material in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

4. I tried hard enough to understand the course material. *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

5. I did not understand the course material because I didn't try hard enough. *

Mark only one oval.

	1	2	3	4	5	6	7	
NOT at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

6. I believe I will receive an excellent grade in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

7. I'm certain that I understood the most difficult material presented in the readings for this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

8. I'm confident that I understood the basic concepts taught in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

9. I'm confident that I understood the most complex material presented by the instructor in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

10. I'm confident that I did an excellent job on the assignments and quizzes in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

11. I expect to do well in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

12. I'm certain that I mastered the skills that were taught in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

13. Considering the difficulty of this course, the instructor, and my skills, I think I was able to do well in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

14. When studying the course, I practiced saying the material to myself over and over. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

15. When studying for this course, I read the course notes and other materials over and over again. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

16. I memorized key words to remind me of important concepts in this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

17. I made a list of important terms for this course and memorized the list. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

18. I used to study in a place where I was able to concentrate on my coursework. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

19. I made good use of my study time for this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

20. I found it hard to stick to the study schedule for this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

21. I had set a regular time studying this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

22. I made sure I kept up with the weekly readings and assignments for this course. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

23. I attended course briefings regularly. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

24. I often found that I didn't spend very much time on this course because of other activities. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

25. I rarely found time to review my notes before an assignment, quiz or exam. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

26. I often felt so lazy or bored when I studied for this course that I used to quit before finishing what I planned to do. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

27. I worked hard to do well in this class even if I didn't like what I was doing. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

28. I gave up studying the difficult part of coursework and only studied the easy parts. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

29. I managed to study until the end even when course materials were dull and uninteresting *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very true of me

30. When I had trouble learning the material in this course, I tried to study the course on my own without seeking help from colleagues or instructor. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

31. I used to ask the instructor to clarify concepts that I didn't understand well. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

32. When I could not understand the material in this course, I asked another student in this course for help. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

33. I tried to identify students in this course whom I could ask for help when it was necessary. *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all true of me	◯	◯	◯	◯	◯	◯	◯	Very true of me

Appendix VII: Post Study Questionnaire

This appendix contains the post-study questionnaire that was used to examine students' perception of the usefulness of the EDM plugin in promoting SRL learning for the experimental group.

Data Science With Python: Post Study Structured Survey

First, thank you for participating and completing the Data Science with Python course. The purpose for this survey is to establish students' perception on the usefulness of the of the intervention messages that were provided by the LMS during the course.

Kindly take your time to respond to the questions in this questionnaire based on your experience during the Data Science with Python course. Respond to the questions and be as honest and accurate as possible. Your participation in this survey is voluntary and the responses will only be used for academic purposes and information you provide will be treated confidentially.

*Required

1. Email *

____

2. Gender *

Mark only one oval.

- ◯ Female
- ◯ Male

3. Level of study *

Mark only one oval.

- ◯ Degree
- ◯ Diploma
- ◯ Other: ______________________________

4. Indicate how often you utilize the following tool(s) to communicate with your colleagues and instructors in class *

Mark only one oval per row.

	Very Often	Often	Sometimes	Rarely	Never
Email	◯	◯	◯	◯	◯
Text messages (SMS)	◯	◯	◯	◯	◯
Social media i.e. WhatsApp/Facebook	◯	◯	◯	◯	◯
Phone calls	◯	◯	◯	◯	◯

5. Did you have any previous experience in online/blended learning through Learning Management System prior to undertaking the Data Science with Python course *

Mark only one oval.

◯ Yes

◯ No

6. Did you have any previous experience in Data Science or Machine learning before the commencement of this course? *

Mark only one oval.

◯ Yes

◯ No

7. Which of the following analytics/statistical tools did you have experience or knowledge in before the commencement of the Data Science with Python Course? Select one or more that apply *

Tick all that apply.

- [ ] SPSS
- [ ] R
- [ ] Python
- [ ] Scala
- [ ] Julia
- [ ] Excel
- [ ] None of the above

Other: [ ] ____________________

8. In your own opinion, what hindered you from actively being involved in the online learning course? Select one or more that apply *

Tick all that apply.

- [ ] Lack of instructor guidance
- [ ] Lack of individualized feedback on your learning habits
- [ ] Lack of peer interaction
- [ ] Lack of interaction with course instructor
- [ ] Lack of adequate internet
- [ ] Lack of adequate learning materials
- [ ] None of the above

Other: [ ] ____________________

9. To what extent would you agree or disagree with the following statements in relation to your experience on the online data science with python course *

Mark only one oval per row.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The LMS provided a platform that made me feel that my individual needs were met by providing individualized intervention messages					
The learning intervention messages provided by the LMS helped me enhance my engagement level					
By providing the summary of learning activities, the LMS helped me improve on the areas I was not doing well					
The intervention messages provided by the LMS helped me improve on poor learning habits					

10. Based on your learning experience, to what extent would you agree or disagree with the following statements in terms of how the LMS helped you to utilize the following strategies *

Mark only one oval per row.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Manage my time well while studying online	◯	◯	◯	◯	◯
Monitor my online learning strategies	◯	◯	◯	◯	◯
Evaluate myself in relation to the progress on learning activities in order to achieve course goals	◯	◯	◯	◯	◯
Seek for help from my colleagues	◯	◯	◯	◯	◯
Seek for help from course instructor	◯	◯	◯	◯	◯
Effort management in terms of knowing the areas that I needed to commit more time and effort	◯	◯	◯	◯	◯
To plan and organize my course learning activities and schedules	◯	◯	◯	◯	◯

11. From your experience during the online learning, indicate the extent to which you agree with the following statements *

Mark only one oval per row.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I was satisfied with the level of support I received through the LMS					
The instructional feedback in terms of the intervention messages I received was timely					
The email notifications that I received helped me track the intervention messages easily					
I was satisfied that my individual learning needs were catered for during the online course					
I was satisfied with the regular reminders that I received how to utilize the LMS features to improve my learning engagement					

12. Briefly describe the best online learning experience that you had during the data science with python course

13. Briefly describe the worst online learning experience that you had during the data science with python course

14. Based on your experience what key areas would you suggest to be included on the interventions that were provided by the LMS in order to improve your online learning experience

Appendix VIII: System Code Segment

This appendix contains part of the code that was used to implement the EDM model through clustering, classification and presentation of SRL interventions to the students via Moodle LMS.

```
// Pass the encoded array to the classifier class
$classifier = new Classifier(__DIR__."/main.py");
$response = $classifier->getClassification($array);
$response = str_replace( array('[',']') , ''  , $response);

//$response = shell_exec("python3.8 main.py ".$array);

//echo $response;
$recordtoinsert = new stdClass();
$recordtoinsert->assignments_viewed = $assign_view;
$recordtoinsert->assignments_submitted = $assign_submit;
$recordtoinsert->quiz_started = $quiz_start;
$recordtoinsert->quiz_submitted = $quiz_submit;
$recordtoinsert->quiz_reviewed = $quiz_review;
$recordtoinsert->quiz_viewed = $quiz_view;
$recordtoinsert->forums_viewed = $forum_view;
$recordtoinsert->chats_viewed = $chat_view;
$recordtoinsert->time_clustered = time();
$recordtoinsert->resources_viewed = $resource_view;
$recordtoinsert->cluster = $response;
$recordtoinsert->userid = $USER->id;

$recordscluster = new stdClass();
$recordscluster->user_id = $USER->id;
$recordscluster->cluster = $response;

$DB->insert_record("local_edm_student_cluster", $recordtoinsert);
$DB->insert_record("local_edm_clusters", $recordscluster);
```

```
global $DB, $USER;

$sql = "SELECT DISTINCT(cluster) FROM {local_edm_student_cluster} WHERE time_clustered < DATE_ADD(now(), INTERVAL 1 WEEK)";
// $sql = "SELECT * FROM {local_edm_clusters} GROUP BY user_id ORDER BY id";

// $params = [
//     'userid' => $USER->id,
// ];

$clusters = $DB->get_records_sql($sql);
$data = json_decode(json_encode($clusters), true);
```

```php
$user_id = $USER->id;
$cluster_id = $cluster;
$message_text = $array[array_rand($array)];
$message_type = "popup";
$email = $USER->email;

// Send email to user
sendEmail($email, $message_text);

$manager = new manager();
$manager->create_message($cluster_id, $user_id, $message_text, $message_type);
```

```php
if ($cluster === 0) {

    // Array of messages based on cluster level
    $array = array(
        "Through the system records, we have noted that you have not logged into the online
    course for nearly a week. Create time and login and participate in the learning
    activities.", #message which is scheduled last.
        "You are not doing well in accessing e-learning system and interacting with peers
    and instructor, you may consider seeking support from peers and instructor by
    reading, posting and replying to chats and forums.",
        "Create time to access LMS and participate in the online learning activities such as
    chats, forums, quizzes, assignments and reading/listening the learning resources
    such as notes and videos.",
        "According to system records, you have not accessed the online course for nearly a
    week. Could you be facing any challenge? Consider accessing LMS and seek
    support from course instructor via message, chat or forum.",
        "Did you know that you can pause reading and think about the previous content
    covered? Try to reflect on what you know with what you are currently reading.",
        "System records indicate that you have spent little or no time on the online course.
    Spend more time studying course materials and video to help you understand the
    course better.",
        "Try to put effort and time on difficulty areas of the course so as to make it easy for
    you to understand.",
        "You have not attempted any learning activity such as chat, forum, quiz and
    assignments. This may lead to poor performance at the end of the course. Create
    time for online learning!",
        "You have not spent time to study course content and videos, time you self and
    spend time studying course materials.",  #message which is scheduled first.
    );
```

```
$output .= '
    <div class="dialog-modal" id="dialog-modal-popupin" style="display: block">
        <div class="dialog-modal-inn">
            <div id="dialog">
                <div class="modal-header">
                    <div class="title">LMS Motivation!</div>
                    <button onclick="hide_popup(event);" class="close-button">&times;</button>
                </div>
                <div class="modal-body">
                <p>' . $messageG . "  " . $Name . ' </p>
                <p>' . $message . '</p>
                <p style="font-size: 12px">All the best, your course lecturer.</p>
                </div>
                <div class="sure-btn">
                    <button onclick="hide_popup(event);">
                        Cancel
                    </button>
                    <button onclick=\'location.href="' . $CFG->wwwroot . '/local/edm/students.php"\'>
                        Go to Dashboard
                    </button>
                </div>
            </div>
        </div>
    </div>
    ';
echo $output;
```

In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from collections import Counter
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split, cross_val_score, StratifiedKFold, Randomi
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.linear_model import LogisticRegression
le = LabelEncoder()
from sklearn import model_selection
from sklearn.neighbors import NearestCentroid
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score, auc
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
from sklearn.preprocessing import StandardScaler
import pickle
from sklearn.metrics import f1_score
import time
from datetime import datetime
from sklearn.metrics import log_loss
```

In [3]:

```
!pip3 freeze > requirements.txt
```

In [22]:

```
df = pd.read_csv('moodle_data.csv')
df.head(2)
```

Out[22]:

	Unnamed: 0	userid	Assignments_viewed	Assignments_submitted	Assignments_created	Assignments_started	A
0	0	2381	9	1	0	0	
1	5	78581	0	0	0	0	

2 rows × 25 columns

In [23]:

```
df = df.drop(columns = ['Unnamed: 0'], axis=1)
```

In [24]:

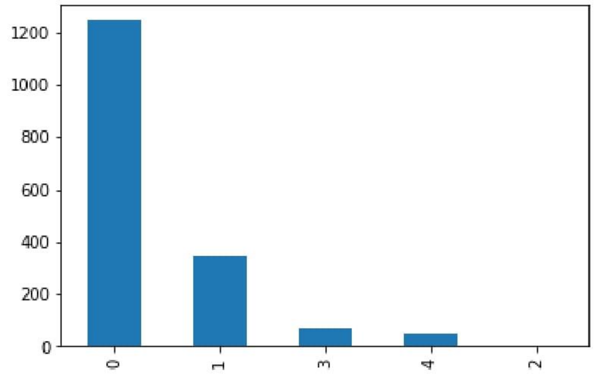
```
df['userid'].nunique()
```

Out[24]: 1710

In [25]:

```
df['Cluster_Category'].value_counts().plot(kind='bar')
```

Out[25]: <AxesSubplot:>



In [26]:

```
#dropping all columns with maximum values of zero
df = df.sample(frac=1).reset_index(drop=True)
df = df[df['Cluster_Category'] != 2]
#try this: may remove
df = df[df['Cluster_Category'] != 3]
to_clean = df.drop(columns=['userid', 'Cluster_Category'], axis=1)
cleaned_data = to_clean[to_clean.columns[to_clean.max() > 0]]
cleaned_data.describe()
```

Out[26]:

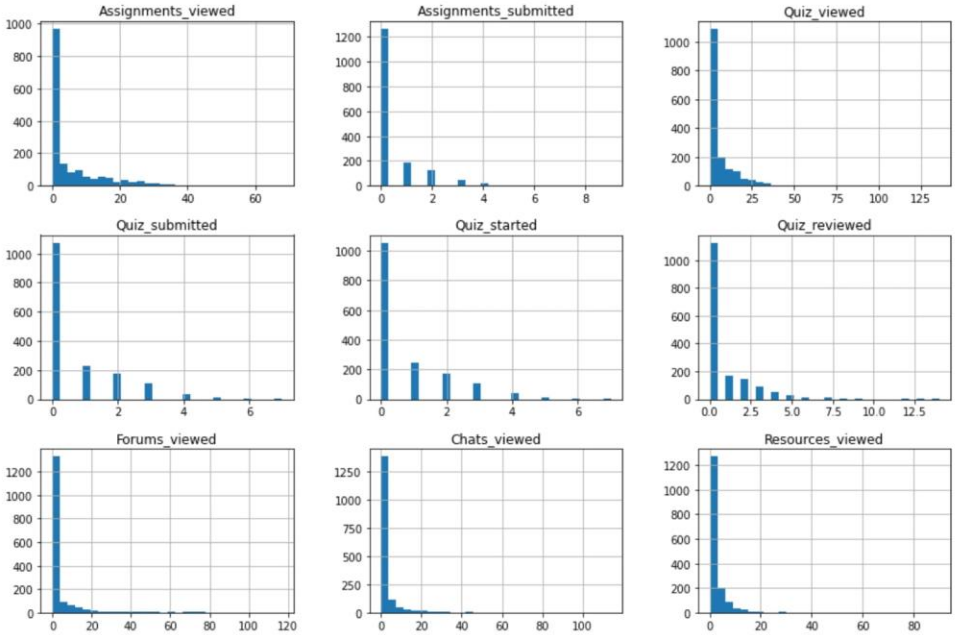
	Assignments_viewed	Assignments_submitted	Quiz_viewed	Quiz_submitted	Quiz_started	Quiz_reviewed
count	1640.000000	1640.000000	1640.000000	1640.000000	1640.000000	1640.000000
mean	5.632927	0.413415	5.556707	0.696951	0.723780	0.826220
std	8.810701	0.908809	9.415717	1.166542	1.189714	1.642828
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000
75%	8.000000	0.000000	8.000000	1.000000	1.000000	1.000000
max	68.000000	9.000000	136.000000	7.000000	7.000000	14.000000

In [27]:

```
cleaned_data.to_csv('plugin_test.csv', index=False)
```

In [28]:

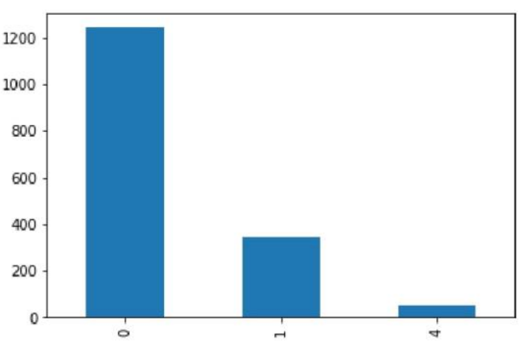
```
cleaned_data.hist(bins=30, figsize=(15, 10));
```



In [29]:

```
df['Cluster_Category'].value_counts().plot(kind='bar')
```

Out[29]: <AxesSubplot:>



In [30]:

```
cleaned_data['userid'] = df['userid']; cleaned_data['Cluster_category'] = df['Cluster_Category'
cleaned_data.head()
```

```
/home/ningi/.local/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/i
ndexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

Out[30]:

	Assignments_viewed	Assignments_submitted	Quiz_viewed	Quiz_submitted	Quiz_started	Quiz_reviewed	Foru
0	0	0	0	0	0	0	
1	1	0	0	0	0	0	
2	1	0	4	1	1	0	

	Assignments_viewed	Assignments_submitted	Quiz_viewed	Quiz_submitted	Quiz_started	Quiz_reviewed	Foru
3	8	0	8	2	2	2	
4	7	0	1	0	0	0	

In [31]:

```
cleaned_data['Cluster_category'] = le.fit_transform(cleaned_data['Cluster_category'])
cleaned_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1640 entries, 0 to 1709
Data columns (total 11 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Assignments_viewed     1640 non-null   int64
 1   Assignments_submitted  1640 non-null   int64
 2   Quiz_viewed            1640 non-null   int64
 3   Quiz_submitted         1640 non-null   int64
 4   Quiz_started           1640 non-null   int64
 5   Quiz_reviewed          1640 non-null   int64
 6   Forums_viewed          1640 non-null   int64
 7   Chats_viewed           1640 non-null   int64
 8   Resources_viewed       1640 non-null   int64
 9   userid                 1640 non-null   int64
 10  Cluster_category       1640 non-null   int64
dtypes: int64(11)
memory usage: 153.8 KB
/home/ningi/.local/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/i
ndexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

In [32]:

```
X = cleaned_data.drop(columns = ['Cluster_category'], axis=1)
y = cleaned_data['Cluster_category']
#checking feature correlation
corr = X.corr()
corr
```

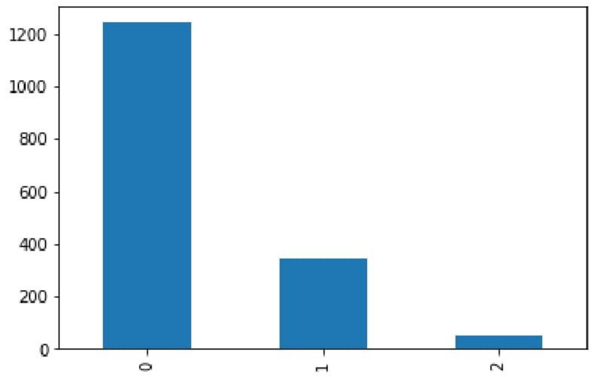
Out[32]:

	Assignments_viewed	Assignments_submitted	Quiz_viewed	Quiz_submitted	Quiz_started
Assignments_viewed	1.000000	0.724243	0.486013	0.481639	0.488565
Assignments_submitted	0.724243	1.000000	0.285885	0.363412	0.362998
Quiz_viewed	0.486013	0.285885	1.000000	0.818313	0.832468
Quiz_submitted	0.481639	0.363412	0.818313	1.000000	0.982868
Quiz_started	0.488565	0.362998	0.832468	0.982868	1.000000
Quiz_reviewed	0.449104	0.313366	0.770318	0.876348	0.867285
Forums_viewed	0.482592	0.227937	0.555738	0.455732	0.449284
Chats_viewed	0.348031	0.185952	0.486342	0.361433	0.360895
Resources_viewed	0.263159	0.108997	0.209024	0.173863	0.176543
userid	0.127107	0.082185	0.029726	-0.009037	-0.007557

In [33]:

```
y.value_counts().plot(kind='bar')
```

Out[33]:
```
<AxesSubplot:>
```



In [34]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
X_tr = X_train.drop(columns=['userid'], axis=1)
X_te = X_test.drop(columns=['userid'], axis=1)
```

In [15]:
```
xgb_params = {
        'min_child_weight': [1, 5, 10],
        'gamma': [0.5, 1, 1.5, 2, 5],
        'subsample': [0.6, 0.8, 1.0],
        'colsample_bytree': [0.6, 0.8, 1.0],
        'max_depth': [3, 4, 5],
        'learning_rate': [0.1, 0.15, 0.05],
        'booster':['gbtree', 'dart']
        }
xgb_model = XGBClassifier(silent=True)
```

In [16]:
```
def timer(start_time=None):
    if not start_time:
        start_time = datetime.now()
        return start_time
    elif start_time:
        thour, temp_sec = divmod((datetime.now() - start_time).total_seconds(), 3600)
        tmin, tsec = divmod(temp_sec, 60)
        print('\n Time taken: %i hours %i minutes and %s seconds.' % (thour, tmin, round(tsec,
```

In [17]:
```
x = X.drop(columns=['userid'], axis=1)
folds = 5
param_comb = 5

skf = StratifiedKFold(n_splits=folds, shuffle = True, random_state = 1001)

random_search = RandomizedSearchCV(xgb_model, param_distributions=xgb_params, n_iter=param_comb
# Here we go
start_time = timer(None) # timing starts from this point for "start_time" variable
random_search.fit(x, y)
timer(start_time) # timing ends here for "start_time" variable
```

```
Fitting 5 folds for each of 5 candidates, totalling 25 fits
```

```
Time taken: 0 hours 0 minutes and 7.46 seconds.
[Parallel(n_jobs=4)]: Done  25 out of  25 | elapsed:    7.0s finished
```

In [18]:

```
print('\nRandom Search Results:')
print(random_search.cv_results_)
print('\n Best Estimator Of RandomSearch:')
print(random_search.best_estimator_)
print('\n Best Hyperparameters from RandomSearch:')
print(random_search.best_params_)
```

```
Random Search Results:
{'mean_fit_time': array([0.53905668, 0.30109706, 0.39205313, 1.16234555, 1.06953955]), 'std_fit
_time': array([0.07610709, 0.03368615, 0.05952015, 0.09765483, 0.21079367]), 'mean_score_time':
array([0.01863976, 0.01153994, 0.01133327, 0.00891399, 0.00997787]), 'std_score_time': array
([0.00685564, 0.00308759, 0.00096347, 0.00132141, 0.00531457]), 'param_subsample': masked_array
(data=[0.8, 0.8, 1.0, 0.8, 1.0],
             mask=[False, False, False, False, False],
       fill_value='?',
            dtype=object), 'param_min_child_weight': masked_array(data=[10, 10, 1, 10, 10],
             mask=[False, False, False, False, False],
       fill_value='?',
            dtype=object), 'param_max_depth': masked_array(data=[5, 4, 4, 5, 5],
             mask=[False, False, False, False, False],
       fill_value='?',
            dtype=object), 'param_learning_rate': masked_array(data=[0.05, 0.15, 0.05, 0.15, 0.
05],
             mask=[False, False, False, False, False],
       fill_value='?',
            dtype=object), 'param_gamma': masked_array(data=[5, 2, 1.5, 5, 5],
             mask=[False, False, False, False, False],
       fill_value='?',
            dtype=object), 'param_colsample_bytree': masked_array(data=[1.0, 0.6, 0.6, 0.8, 0.
8],
             mask=[False, False, False, False, False],
       fill_value='?',
            dtype=object), 'param_booster': masked_array(data=['gbtree', 'gbtree', 'gbtree', 'd
art', 'dart'],
             mask=[False, False, False, False, False],
       fill_value='?',
            dtype=object), 'params': [{'subsample': 0.8, 'min_child_weight': 10, 'max_depth':
5, 'learning_rate': 0.05, 'gamma': 5, 'colsample_bytree': 1.0, 'booster': 'gbtree'}, {'subsampl
e': 0.8, 'min_child_weight': 10, 'max_depth': 4, 'learning_rate': 0.15, 'gamma': 2, 'colsample_
bytree': 0.6, 'booster': 'gbtree'}, {'subsample': 1.0, 'min_child_weight': 1, 'max_depth': 4,
'learning_rate': 0.05, 'gamma': 1.5, 'colsample_bytree': 0.6, 'booster': 'gbtree'}, {'subsampl
e': 0.8, 'min_child_weight': 10, 'max_depth': 5, 'learning_rate': 0.15, 'gamma': 5, 'colsample_
bytree': 0.8, 'booster': 'dart'}, {'subsample': 1.0, 'min_child_weight': 10, 'max_depth': 5, 'l
earning_rate': 0.05, 'gamma': 5, 'colsample_bytree': 0.8, 'booster': 'dart'}], 'split0_test_sco
re': array([-0.12317959, -0.09761396, -0.11615561, -0.11229373, -0.12569799]), 'split1_test_sco
re': array([-0.12633984, -0.0886998 , -0.10520404, -0.1084958 , -0.12148519]), 'split2_test_sco
re': array([-0.13102929, -0.09131861, -0.10916091, -0.11689454, -0.13163195]), 'split3_test_sco
re': array([-0.14432392, -0.12049139, -0.12764581, -0.13778653, -0.14594004]), 'split4_test_sco
re': array([-0.14571992, -0.109535  , -0.1245791 , -0.13625992, -0.14423412]), 'mean_test_scor
e': array([-0.13411851, -0.10153175, -0.11654909, -0.1223461 , -0.13379786]), 'std_test_score':
array([0.00925693, 0.0118982 , 0.00861474, 0.01228496, 0.00978007]), 'rank_test_score': array
([5, 1, 2, 3, 4], dtype=int32)}

 Best Estimator Of RandomSearch:
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=0.6, gamma=2, gpu_id=-1,
              importance_type='gain', interaction_constraints='',
              learning_rate=0.15, max_delta_step=0, max_depth=4,
              min_child_weight=10, missing=nan, monotone_constraints='()',
              n_estimators=100, n_jobs=0, num_parallel_tree=1,
              objective='multi:softprob', random_state=0, reg_alpha=0,
              reg_lambda=1, scale_pos_weight=None, silent=True, subsample=0.8,
              tree_method='exact', validate_parameters=1, verbosity=None)

 Best Hyperparameters from RandomSearch:
{'subsample': 0.8, 'min_child_weight': 10, 'max_depth': 4, 'learning_rate': 0.15, 'gamma': 2,
'colsample_bytree': 0.6, 'booster': 'gbtree'}
```

In [19]:
```
best_params = {'subsample': 0.8, 'min_child_weight': 10, 'max_depth': 4,
                          'learning_rate': 0.15, 'gamma': 2, 'colsample_bytree': 0.6, 'booster': 'g
xgb_model = XGBClassifier(**best_params)
```

In [20]:
```
xgb_model.fit(X_tr, y_train)
predictions = xgb_model.predict(X_te)
```

In [21]:
```
print(confusion_matrix(y_test,predictions))
print(classification_report(y_test,predictions))
```

```
[[312   4   0   0]
 [  5  76   1   0]
 [  0   1  14   0]
 [  0   3   3   9]]
              precision    recall  f1-score   support

           0       0.98      0.99      0.99       316
           1       0.90      0.93      0.92        82
           2       0.78      0.93      0.85        15
           3       1.00      0.60      0.75        15

    accuracy                           0.96       428
   macro avg       0.92      0.86      0.87       428
weighted avg       0.96      0.96      0.96       428
```

In [22]:
```
SEED= 42
#Models Comparison
models = []
models.append(('LR', LogisticRegression()))
models.append(('CART', DecisionTreeClassifier()))
models.append(('RF', RandomForestClassifier()))
models.append(('XGB', XGBClassifier()))
# evaluate each model in turn
results = []
names = []
scoring = 'accuracy'
for name, model in models:
    kfold = KFold(n_splits=10, random_state=SEED)
    cv_results = cross_val_score(model, x, y, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
```

In [35]:

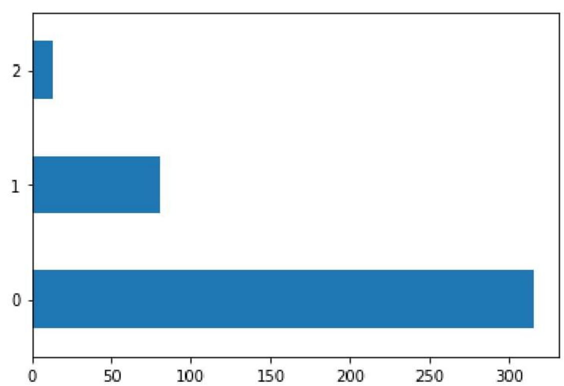
```
log_reg = LogisticRegression()
log_reg.fit(X_tr, y_train)
predictions_lr = log_reg.predict(X_te)
confusion_matrix(predictions_lr, y_test)
print(classification_report(predictions_lr, y_test))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       316
           1       1.00      0.99      0.99        81
           2       1.00      1.00      1.00        13

    accuracy                           1.00       410
   macro avg       1.00      1.00      1.00       410
weighted avg       1.00      1.00      1.00       410

/usr/local/lib/python3.6/dist-packages/sklearn/linear_model/_logistic.py:764: ConvergenceWarnin
g: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
```

In [36]:

```
X_te['predictions'] = predictions_lr
X_te['predictions'].value_counts().plot(kind='barh')
```

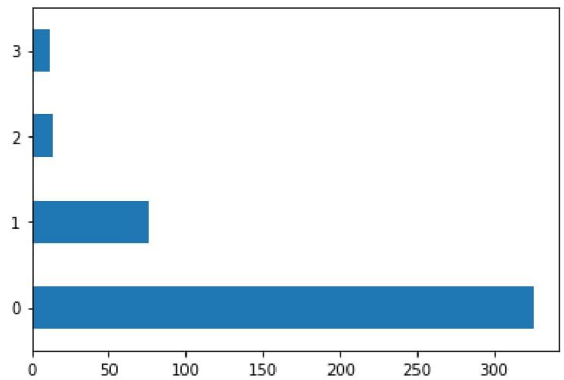
Out[36]: <AxesSubplot:>



In [21]:

```
y_test.value_counts().plot(kind='barh')
```

Out[21]: <AxesSubplot:>



In [24]:

```
filename = 'log_reg_model.sav'
pickle.dump(log_reg, open(filename, 'wb'))
```

In [ ]: