UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

INTELLIGENT SUPPORT IN GROUP WORK IN ONLINE COLLABORATIVE LEARNING ENVIRONMENT

ELIZAPHAN MUURO MAINA

P80/83902/2012

SUPERVISORS:

PROF. PETER WAGACHA

DR. ROBERT OBOKO

MARCH, 2015

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE OF THE UNIVERSITY OF NAIROBI

DECLARATION

This thesis is my original work and has not been presented for examination in any other University. No part of it may be reproduced without the prior permission of the author or the University of Nairobi

Signature: .... Date: ....06/03/2015....

Elizaphan Muuro Maina

This thesis has been submitted for examination with our approval as University supervisors:

Signature: .... Date: ....9th March 2015....

Prof. Peter Wagacha

School of Computing and Informatics,

University of Nairobi, Kenya

Signature: .... Date: ....10/3/2015....

Dr. Robert Oboko

School of Computing and Informatics,

University of Nairobi, Kenya

DEDICATION

To my dear wife and family members and colleagues who supported me throughout the completion of this project.

ACKNOWLEDGEMENT

First, I would like to thank almighty God for His power and grace which have enabled me to complete this research journey. I also wish to thank the National Commission for Science, Technology and Innovation (NACOSTI), Kenya for funding this research.

Special thanks to my supervisors Prof. Peter W. Wagacha and Dr. Robert O. Oboko for their endless support throughout the various stages of this research. Your insights and suggestions have always been on target and timely. Your valuable guidance on how to publish greatly improved my research ability and confidence to publish my research work. I am indebted to your effort, time, moral support, and the valuable comments you provided throughout the entire research process.

I would also like to thank the students from Kenyatta University (KU), Jomo Kenyatta University of Science and Technology (JKUAT), United State International University, Kenya (USIU-Kenya) and Australian Study Institute, Kenya (AUSI-Kenya) who participated in this research project. Thanks to Prof. Jimmy Macharia of USIU, Dr. Karimi of KU, Mr. Harrison of AUSI and Mr. Simon Maina of JKUAT for their support during data collection in these universities. Thanks to Prof. John Kihoro for his valuable suggestions regarding the Moodle e-learning platform. Thanks to eBay's info solutions for their technical support in Moodle installation and hosting. I would also like to thank my colleagues in the department of Computing and Information Technology at KU and my fellow graduate students for their moral support.

A very special thanks to my wife, Salome for the absolute moral support, encouragement, and love she provided from the start to the end of this project. Thank too to my children (James and Army) for your patience when dad was away completing this project. Finally, I would like to thank my parents, brothers and sisters for their moral support and encouragement throughout this writing process.

ABSTRACT

The increased demand for higher education has made online learning popular and appealing to many stakeholders including working staff and students. Though online learning has gained popularity, it is still being criticized for being a faceless medium that does little to support social interaction. Social constructivist argue that knowledge is constructed through social activities and therefore, knowledge developed using collaboration is more than what can be achieved by an individual alone. Online learning, if supported by a good collaborative strategy like discussion forums, can be at par with social constructivist view of learning in terms of learning achievement. Learning Management Systems such as Modular Object-Oriented Dynamic Learning Environment (Moodle) supports online tools that include discussion forums, chat rooms, e-mails, newsgroups, workshops, etc. These tools provide new opportunities for students to collaborate online and construct knowledge through peer learning. Despite the pedagogical advantages of collaborative learning, online learners can perceive the collaborative learning process as challenging. Although a number of challenges have been mentioned in the literature, they do not have empirically grounded evidence and therefore overcoming these challenges remains an issue. In light of this and the increasing demand for online learning, this research aimed at investigating the current status of online collaborative learning in Higher Learning Institutions in Kenya, identify perceived challenges, and explore strategies for improving online collaborative learning through intelligent support techniques such as machine learning.

To that end, this research was designed using a Multi-Methodological approach in order to develop and validate a prototype which provided a novel approach through intelligent support techniques for group formation based on students' collaboration competence level and a platform to provide immediate feedback in Moodle. The first part of the methodology was a cross-sectional survey which was used to carry out a pre-study to investigate the current status of online collaborative learning and students' perceived challenges in an online collaborative learning environment in Higher Learning Institutions in Kenya. This pre-study informed the system development and the validation processes for the prototype. The second part of methodology was the system development methodology which guided the development of the prototype. The final part of the methodology an experimental design that was carried out to evaluate the effectiveness of the intelligent module on the formation of diverse groups and the impact of the group formation method on group performance in an online collaborative learning environment. In this study three groups were used, where in group one students were assigned into groups using grade point average scores , in group two students were assigned into groups using intelligent grouping algorithm and in group three students were assigned into groups using the random method.

The novel approach for group formation using machine learning techniques was found capable of forming heterogeneous groups that tended to perform effectively and efficiently at the same level as the random method and the grade point average method. There was no statistical significance in differences associated with the three methods of group formation and performance in a group task. Furthermore, all the three groups had similar positive learning experiences and had few common challenges. With the understanding that random assignment method only increases the likelihood of heterogeneity in the group and grade point average method involves the instructors and it is not dynamic, our proposed intelligent grouping algorithm has the advantage of guaranteeing heterogeneity based on learner's collaboration competence level, dynamism in grouping students and less instructor involvement. Due to these advantages, instructors are more likely to adopt our intelligent grouping technique. Further studies should also be conducted to compare the intelligent grouping with other group assignment methods which were not studied in this study.

Keywords: Social Constructivist, Online Collaborative Learning, Intelligent Support, Group Formation, Collaboration Competence Level, Learner Management System .

ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
ARFF	Attribute-Relation File Format
AUSI	Australia Studies Institute
CAT	Continuous Assessment Test
CMC	Computer-Mediated Communication
CSCW	Computer Supported Collaborative Work
DB	Database
EM	Expectation Maximization
GPA	Grade Points Average
HLIs	Higher Learning Institutions
IBL	Instance Based Learning
ICT	Information and Communication Technology
ITS	Intelligent Tutoring Systems
JKUAT	Jomo Kenyatta University of Science and Technology
KU	Kenyatta University

LMS	Learning Management System
MKO	More Knowledgeable Other
ML	Machine Learning
Moodle	Modular Object-Oriented Dynamic Learning Environment
ODeL	Open and Distance e-Learning
RDBMS	Relational Database Management System
SOM	Self-Organizing Maps
UoN	University of Nairobi
USIU	United State International University
Weka	Waikato Environment for Knowledge Analysis
ZPD	Zone of Proximal Development

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER ONE

INTRODUCTION

1.1 Background

The increased demand for higher education has altered the traditional nature of formal education commonly referred as 'classroom environment' to an online environment where students undertake programs and courses on the internet (Roberts, 2004). Several institutions of higher learning are struggling with resources such as space to meet the increased demand for education and online education provides the promise. Thus, the emergency of collaborative learning in online environments is in response to the rapid increase in demand for online education. The online environment fulfills the need for social interaction from a diverse range of backgrounds, different learning abilities and the desire to study on their own preferred times and places.

According to Roschelle and Teasley (1995) collaboration is a process by which individuals negotiate and share meanings relevant to the problem solving task at hand. In collaboration, there is always a continued attempt to construct and maintain a shared conception of a problem. Dillenbourg (1999) defines collaborative learning as a situation whereby two or more people learn or attempt to learn something together. The situation is termed "collaborative" if the peers are more or less at the same level, can perform the same actions, have common goal and work together.

In social constructivist theory of learning, social interaction plays a fundamental role in the process of cognitive development which can be assessed within the Zone of Proximal Development (ZPD) (Vygotsky, 1978). In this ZPD, Vygotsky (1978) claims that learner's performance can be optimized by collaborating with more capable peers. This constructivism theory of learning has been adopted in Higher Learning Institutions (HLIs) where students are engaged in discussion by tutorial fellows. These tutorials give the learners a chance to collaborate face to face, critique one another, share knowledge and compare new concepts with one another. Moreover, social constructivists' scholars argue that knowledge is constructed through social activities and therefore, knowledge developed through collaborative learning is more than what can be achieved by an individual alone. Despite all these pedagogical advantages of collaborative learning, the literature has reported a number of challenges encountered by students in group tasks (Roberts and McInnerney, 2007; Capdeferro and Romero, 2012). However, in the Kenyan context, results from various studies in the literature reviewed are not empirically grounded.

The use of artificial intelligent mechanisms to support the collaborative learning process has also provided new mechanisms which can improve learners' interaction in an online collaborative learning environment. For example, recent research has tested the application of Machine learning (ML) techniques in different aspects of collaborative learning (Anaya and Boticario, 2009b, 2010, 2011a). This provides new opportunities for managing online collaborative learning with little intervention by the instructor. Even so, research in the use of ML algorithms to improve online collaborative learning has not received sufficient attention as some aspects of collaboration like group orientation through intelligent mechanisms is yet to be implemented in e-learning platforms.

1.2 Statement of the Problem

With the increase in demand for higher education, e-learning environment has gained popularity. Majority of HLIs have invested in e-learning systems with the hope of saving costs associated with traditional learning systems and to capitalize on the globalized demand for higher education. These e-learning systems do provide collaboration tools like discussion forums which can allow social interaction and learning between peers. However, this social interaction lacks the aspect of face to face interactions typical of classroom environments. This gives online learning a major disadvantage even when its demand continues to rise. Consequently, online collaborative learning remains more challenging than face to face learning prompting the need to carry out more empirical research to identify the key challenges and provide mechanisms to address them. Additionally, further research is required to identify the pedagogical issues in collaborative learning are not fully understood in order to assist instructors to understand the pedagogical issues in group work as well as helping students to reflect on their peer learning. Although a number of challenges have been mentioned in the literature, considerable diversity exists among countries due to diversity in infrastructure support for e-learning and learners' background. This gave the impetus to investigate the current status on online collaborative learning and the perceived challenges by learners in HLIs in Kenya.

In recent years, ML techniques have been applied to support the collaborative learning process and improve learners' interaction in e-discussions (McLaren et al., 2010; Anaya and Boticario, 2009b, 2010, 2011a). Moreover, recent research including Anaya and Boticario (2009b, 2010, 2011a) has revealed that ML techniques can be applied to analyze students interaction in group work and rank learners according to their collaboration level. This helps learners and tutors evaluate the collaborative work and identify possible problems as they arise. However, these studies do not address the aspect of group formation which posses a positive impact in group performance. Without appropriate support in group formation, students tend to form groups which are more social but ignore aspects of collaboration competence level. For example, self created groups tend to be more associated with demographic characteristics while randomly created groups could be homogenous rather than heterogeneous in terms of individual capabilities. Moreover, current research does not suggest an algorithm which can group students based on their collaboration competence level. This aspect also gave impetus to explore group formation methods further, alongside feedback to be provided to realize intelligent support for online collaborative learning.

1.3 Purpose of Research

The purpose of this research was to investigate and experiment on ways of improving learner performance in online collaborative learning using intelligent grouping and provision of feedback in an online collaborative learning environment.

1.4 Research Questions

The research addresses the following research questions:

1. What is the current status in terms of collaboration tools, group orientation and collaboration activities on online collaborative learning in HLIs in Nairobi, Kenya?

2. What are the components of online collaborative learning which learners perceive as challenging in HLIs in Nairobi, Kenya?

3. Which group of learners amongst the intelligently grouped, randomly grouped and instructor grouped using Grade Point Average (GPA) scores, collaborates more effectively and performs better in an online group task?

4. What is the association between grouping method used and group outcomes in terms of: a) students' learning experiences; b) perceived problems; c) group leadership satisfaction and; d) group task satisfaction?

5. What are the students' perceived benefits and challenges of online collaborative learning?

1.5 Thesis Outline

Chapter One provides a brief introduction to the background of the study, the problem statement, research purpose, research questions and structure of the thesis.

Chapter Two presents literature review of collaborative learning environment, challenges and problems associated with online collaborative learning in HLIs, group formation techniques, use of machine learning techniques in e-learning and application of clustering techniques in collaborative learning. This chapter also reviews the relevant theories and pedagogical issues in an online collaborative learning environment and identifies the gaps in the existing literature which informs this research and provides a conceptual framework for this research.

Chapter Three presents the system development methodology. A multi-methodological approach that was used in system development in this study is described. Firstly, pre-study survey methodology is presented and the results are also summarized in order to provide the background information which was required to develop and validate the system. Secondly, the methodology for prototype development is discussed. This methodology discusses how an intelligent grouping algorithm and a feedback platform based on learner's collaboration competence level was developed and integrated into Moodle in order to come up with the required prototype.

Chapter Four presents the second part of the methodology which was used to validate the prototype. An experimental design methodology is presented which discusses how the system was evaluated using two control groups. The methodology provides summative and formative evaluation techniques for the prototype in a real world online collaborative learning environment.

Chapter Five presents the findings from the experimental study. Comprehensive discussions on the findings are also presented.

Chapter Six concludes the thesis, highlighting the limitations and summarizing the contributions of this research. This chapter also suggests possible future research work.

1.6 Language Conventions

The following set of words in this study have been used interchangeably depending on the context:

Learner - Student

Group orientation - Group formation - Group member assignment

Group task - Group work

Higher Learning Institutions (HLIs)- High Education Institutions (HEIs) - Universities

Instructor - Teacher - Lecturer

His has been used to represent both sexes.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The increased demand for higher education has made online learning popular and appealing to many stakeholders including working staff and students. Although online learning has gained popularity, it still criticized as being a faceless medium that does little to support social interaction. Social constructivism scholars argue that knowledge is constructed through social activities and therefore, knowledge developed through collaborative learning is more than what can be achieved by an individual alone. Researchers have suggested that learning is more effective if peers collaborate and share ideas when solving a task as a group rather than as individuals (Johnson and Johnson, 1989). They also suggest that construction and synthesis of knowledge through group work outperforms individual learning (Brindley et al., 2009; Moller, 1998).

This constructivist theory of learning has been adopted in HLIs in Kenya where students are engaged in discussions by tutorial fellows. The tutorials give the learners a chance to discuss face to face, critique one another, share knowledge and compare new concepts with one another. Similarly, by introducing e-discussion forums in an online learning environment, it is possible to have social interaction and learning between peers. However, social interaction experienced in an online learning environment lacks the face to face interactions typical in a classroom environment. This gives online learning a major disadvantage even though its demand continues to rise.

According to Vygotsky (1978), group work is more fruitful when learners discuss with experts or more knowledgeable peers because what an individual does jointly with others can be incorporated into his individual problem solving process. Thus, there is a need to constitute a heterogeneous group in group work which constitutes students with different learning competencies. Studies have been conducted to establish the effect of group membership in group work (North et al., 2000; Schellenberg, 1959; Fraenkel et al., 2012), but with limited focus on how different group orientation techniques affect group performance in a groups task. Moreover, intelligent grouping techniques which require little intervention by the instructor are yet to be explored in Moodle.

Currently, HLIs have embraced the use of Learning Management System (LMS) like Moodle and Blackboard. These LMSs are used as platforms to deliver online learning where instructors post e-learning materials and assignments. Most of the learning tools available on these LMS such as chats, forums, wikis, quizzes and workshops are not fully utilized by instructors (Muuro et al., 2014b). Although both Moodle and Blackboard are LMSs, Moodle is open source software which supports both individual learning and group learning. These two elements make it to be widely used in HLIs. On the other hand, Blackboard is commercial software which supports individual learning and provides limited support on group work. The availability of Moodle as open source software makes it easier to develop plug-ins which can improve e-learning process. Lack of enough instructors to support teaching and learning in these LMSs has prompted researchers to develop Intelligent Tutoring Systems (ITS) which can be integrated into these systems. In addition, artificial intelligent techniques like ML techniques have been applied to analyze learner interaction data and develop mechanisms which can improve online learning in e-learning platforms (Anaya and Boticario, 2009b, 2010, 2011a).

The rest of the chapter is organized as follows:

Section 2.2 introduces the concept of e-learning and related terms. Section 2.3 introduces the concept of online learning. Section 2.4 discusses collaborative learning. Section 2.5 discusses online collaborative learning. Section 2.6 discusses tools which are used to analyze collaboration in online learning environments. Section 2.7 discusses group formation techniques in online collaborative learning and identifies the merits and demerits of different group formation methods when doing a group task. Section 2.8

discusses perceived challenges of online collaborative learning. Section 2.9 discusses the practice of blended learning in HLIs in Kenya. Section 2.10 introduces ML and their application in e-learning. Section 2.11 discusses the use of clustering algorithms in e-learning and two clustering algorithms i.e. SKmeans and EM are discussed. Section 2.12 discusses the application of ML techniques in collaborative learning. Section 2.13 describes the conceptual framework for the experimental design methodology and lastly section 2.14 gives the summary of the chapter.

2.2 e-Learning

E-Learning also referred as electronic learning has been defined in many ways by different authors and therefore it can be a difficult term to define. Some authors define e-learning as a type of learning whereby learning materials are accessed through technological tools that are web-based, web-distributed or web-capable (Nichols, 2003). Other authors include other technological aspects like audio and video tapes, satellite broadcast and interactive TV (Clark, 2002). Even so, these technological aspects are not enough to define e-learning as effective learning must consider other aspects of the knowledge building process which enable the learner to transform experience into individual's knowledge. Therefore, this requires an e-learning definition to consider constructivist learning models which advocate the use of collaborative learning (Vygotsky, 1978). Furthermore, some authors believe that some level of interactivity should be included in describing the learning experience (Ellis, 2004). Terms such as online learning, web-based learning, web-based training and distance learning have also been used by some authors to refer to e-learning (Dringus and Cohen, 2005; Khan, 2001; Triacca et al., 2004; Wagner, 2001). Consequently, it is evident that there is some uncertainty on the extent to which we should define e-learning in order to include all these characteristics. Clearly, any form of definition should not ignore those technological aspects and must include learning models which are applicable in providing learning opportunities responsible for knowledge construction.

For the purpose of this study, the term e-learning was used to encompass web-based education or virtual learning environments, and also where the learning process can occur electronically anytime and anywhere via the internet or intranets (Wang et al., 2011). Therefore, e-learning changes the traditional form of learning which is class-based to a more flexible learning process with considerable benefits such as: convenient access to learning materials, interconnectivity of learners anywhere and anytime, real-time content update and just-in-time training (Wang et al., 2011). E-learning also provides a chance to learners to have self-paced learning which is self-directed (Abdelaziz et al., 2011). Despite these advantages, e-learning can also be challenging since learners need to have computer skills and also have access to computers and internet. Therefore, problems such as lack of computer skills and slow internet connectivity could frustrate learners who might eventually give up (Art and Lisa, 2008). Instructors are also required to be trained in e-learning pedagogies in order to deliver learning materials more effectively in electronic modes. In addition, more resources such as extra time and skilled instructors who are dedicated to using technology-enhanced training are needed.

With emerging technologies such as Web 2.0 (wikis, blogs, social networks, workshops, forums, etc), today's e-learning can allow learner-centered learning which is self-paced and allows individual learner development and at the same time creates a community of knowledge between learners themselves and also between learners and teachers (Albidewi and Tulb, 2014). Also, emerging technologies such as open source web based e-learning environments coupled with higher internet speed in mobile devices have resulted in mobile learning thus enabling more learners to access e-learning services anywhere, anytime. These technologies have made it possible to build open source e-learning platforms such as Moodle which provides face-to-face instructional experience. These e-learning platforms provides configurable infrastructure that integrate learning materials and services to a single solution hence, creating a learning platform which can effectively deliver educational content (Abdelaziz et al., 2011). Even with these technologies, designing effective e-learning materials still remains a challenge (Hutchings et al., 2007). Further, well designed courses must promote active learning, peer learning, and interactivity between teacher, learners and materials must be ensured. Therefore, there is need to conduct more research on pedagogical issues in e-learning given the modern technology trends and the promising future of e-learning.

2.3 Online Learning

Just like e-learning, online learning has also been defined in relation to technological characteristics making it difficult to distinguish from e-learning. Similarly, some authors conclude that online learning is synonymous to e-learning (Dringus and Cohen, 2005; Khan, 2001). Others (Benson, 2002; Conrad, 2002) relate it with distance education where technology is used to provide access to learning to those who are geographically distanced and are unable to attend face-to-face physical classrooms. Others have provided variance by describing online learning as 'purely' online (Oblinger and Oblinger, 2005) or hybrid/blended learning (Olapiriyakul and Scher, 2006). Therefore, online learning becomes a mode of learning where the instructor utilizes technology with intranet or internet-based tools to carry out learning activities such as instruction delivery, assessment and communication and achieve the same learning objective as it could be achieved in face-to-face physical classroom. This makes online learning to become an alternative to classroom learning. Due the increased demand in education, online learning has gained popularity and therefore teaching online is no longer a new event. Use of online technologies to supplement face-to face instruction has yielded blended learning (Ganzel, 2001; Laster, 2003; Mantyla, 2001) which has changed the traditional learning environment.

Blended learning has become the most popular mode of online learning rather than pure online learning which completely lacks the human aspect in delivery and assessment. The lack of human aspect in pure online learning makes it more challenging in terms of addressing ethical and sociological issues in human learning than blended learning.

This calls for further research in technology-enhanced learning which is able to lend itself to the human element in order to relate to the human psyche and philosophy (Cunningham et al., 2014). This study therefore, has adopted blended learning which has proven to be effective and has been implemented in various universities in developing countries like Kenya to provide education to distance students and to complement face to face learning due to the increased enrollment despite challenges such as slow internet (Kashorda and Waema, 2014) and lack of enough and skilled instructors (Nyerere et al., 2012).

As online technologies in e-learning systems gain popularity, problems of attrition and motivation among participants cannot be ignored (Harasim et al., 1995; McConnell, 2000). Hence more focus is needed in the overall design of the learning environment, the specific instructional design (Harasim et al., 1995; Gunawardena et al., 1997) and the experiences the participants have in online learning programs (Sage, 2000). In response to these issues, researches have been conducted which encourage the design of an online learning environment with active participation among learners, based on real-life problems, which includes learner activities grounded in the learner's life context and experiences (Barrows, 1994; Koschmann, 1996; Jonassen, 1997; Oboko and Wagacha, 2012) and creates a sense of community (Palloff and Pratt, 1999). The next section examines collaborative learning which has been used to provide active participation based on constructivist view of learning.

2.4 Collaborative Learning

Collaborative learning is an overloaded term with different meanings offered by different scholars. In this study, Dillenbourg (1999) definition has been adopted where collaborative learning is defined as situation in which two or more people learn or attempt to learn something together. The situation is termed "Collaborative" if peers are more or less at the same level, can perform the same actions, have a common goal and work together. Collaborative learning techniques in different learning modes have been experimented and found to be successful as early as the late 18th century (Gaillet, 1994). Collaborative learning technique involves the use of group work to perform a task. In the learning pedagogy, teachers are always encouraged to assign group work offering the students the freedom to learn from one another. The idea of group work in learning finds its root from the Russian psychologist, Vygotsky (1978), who explored the causal relationship that exists between social interaction and individual learning, providing a foundation for the social constructivist theory of learning. Vygotsky's (1978) theory gives three major themes:

1. Social interaction plays a fundamental role in the process of cognitive development. In contrast to Piaget (1929), understanding of child development in which development necessarily precedes learning, Vygotsky (1978) felt that social learning precedes development. He states: "Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological)" (p57).

2. The More Knowledgeable Other (MKO) concept. The MKO refers to anyone who has a better understanding or a higher ability level than the learner, with respect to a particular task, process, or concept. The MKO is normally thought of as being a teacher, coach, or older adult, but the MKO could also be peers, a younger person, or even computers.

3. The Zone of Proximal Development (ZPD) concept. The ZPD is the distance between a student's ability to perform a task under guidance from an adult/teacher and/or with peer collaboration, and the student's ability to solve the problem independently. In regard to this theory, learning occurs in this zone.

Vygotsky's theory advocates learning contexts in which students play an active role in learning, shifting the role of the teacher from being the centre of knowledge into a facilitator who will help the students form groups and facilitate collaborative learning. Learning therefore becomes a reciprocal experience for the students and the teacher.

Even though there are different theories of learning, a number of them have stressed the importance of interaction and in particular the use of group work (Jonassen, 1997). Use of group work in collaborative learning not only gives academic benefit but also social and psychological benefits which can stimulate critical thinking (Panitz, 1997). Groups which share the same or common education goals can have positive collaborative learning since students are able to share their doubts, comments and questions within the group (Olguin et al., 2000). Such collaboration justifies group or individual action to each other, and this will often lead to great understanding of the information being shared (Dillenbourg and Schneider, 1995). By using group work, learning is shifted from teacher control to the student peer groups, consequently helping learners to acknowledge their dissent, disagreements and share their doubts (Bruffee, 1999).

Tinzmann et al. (1990) suggest four typical characteristics of collaboration in classroom environment:

1. Shared knowledge between teachers and students where the teacher is not only the giver but incorporates student input allowing students to share experiences or knowledge.

2. Shared authority between teachers and students, where the teacher shares the setting of the goal within a topic with the students, allowing the students to complete an assignment in a manner of their choice.

3. Teacher as a mediator: teacher's role is to encourage the students to teach each other.

4. Heterogeneous groupings of students: this encourages all students to respect and appreciate the contributions made by all members of the class, no matter the content.

Therefore, the goal for collaborative learning can be understood as the creation of learning situations in which productive interaction between learners should occur (Ronteltap and Eurelings, 2002). Unlike individual learning where learners are just consumers of knowledge, in collaborative learning, learners are considered co-constructors of knowledge (Bruffee, 1999). Entirely collaborative learning can be viewed as having three components namely: collaboration, communication and social context (Brandon and Hollingshead, 1999). Although researchers have examined the three components separately (Talavera and Gaudioso, 2004), it is important to examine them holistically.

2.5 Online Collaborative Learning

Online collaboration is recognized by Palloff and Pratt (2004) as an educational approach that is based on the constructivist view of learning requiring learners and instructors to work together when solving problems, completing tasks, or creating products. Creating online learning communities rich in collaborative learning tasks has been pointed out as of major benefit to adults who can share work related experiences around the globe (Bonk and Kim, 1998). In the past, collaborative learning has been restricted to the classroom environment because of the logistical difficulties in distance learning environment (Kimball, 2001). However, the introduction of internet technologies and other online tools offers new opportunities for student collaboration in an online environment as well as posing new challenges for teachers supporting group work (Bonk et al., 1998; Palloff and Pratt, 1999).

To promote online collaborative learning, features which support social interaction such as text-based and computer-mediated interaction, many-to-many communication, time and place independence and hypermedia must be part of the online environment (Warschauer, 1997). However, with Web 2.0 technologies, these features are no longer a challenge. The Web 2.0 tools have changed the face of online collaborative learning by providing new opportunities for social interaction which have expanded the learners' role from being passive recipients of knowledge to active participants in knowledge construction (Brown and Adler, 2008). Anderson (2009) identifies some key features in harnessing the power of the crowd, individual production and user generated content, data on an epic scale, architecture of participation, network effects and openness which can be used to maximize the impact of Web 2.0 technologies on promoting online collaborative learning.

Web 2.0 technologies provide various benefits related to online collaborative learning. These benefits are summarized in Table 2.1.

Table 2.1: Summary of advantages of Web 2.0 related to online collaborative learning

Advantage	Source
• Generate deeper level of knowledge • Promotes initiatives, creativity, and critical thinking • Create common goals and form the foundation for a learning community • Address different learning styles and the use of multiple skills • Address issues related to learners culture	(Palloff and Pratt, 2004)
• Enabling under-represented population to contribute in equal proportion as their peers	(Anderson and Lin, 2009)
• Supporting many-to-many and time-and-place independent interactions	(Warschauer, 1997)
• Preparing learners for life-long learning activities	(Weinberger et al., 2005)
• Providing more potential for competence development that empower the learner to become self-guided and self-organized individuals	(Ehlers, 2008)
• Encourages learners to engage in cognitive restructuring or elaboration of learning material	(Slavin, 1996)

Group definition, assigning students to groups and establishment of communication sessions (either synchronous or asynchronous) is of paramount importance for the success of online collaborative learning (Olguin et al., 2000). The same researchers argue that collaboration is more effective if the groups are composed of learners with same interest. Lack of collaborative learning strategies in an online learning system is a major disadvantage to learners as the social relationship aspect is lost (Hiltz, 1998).

Thus, strengthening interactions and group activities is central to the facilitation of online learning.

The use of discussion forums for online learning gives the learners a chance to collaborate online, critique one another, share knowledge and compare new concepts with one another. Discussion forums create a platform where learners can learn on their own with the opportunity of sharing experiences and constructing knowledge based on their cognitive level (Corich and Hunt, 2004). Garrison (1993) on his discussion for constructivist approach to learning argued that "Learners attempt to interpret, clarify and validate their understanding through constructed dialogue and negation" (p. 102). When discussion forums are managed well, they constitute a major tool for supporting e-learning as they encourage learners to share knowledge and build new ideas from shared concepts (Garrison, 1993). Harasim et al. (1995) also stipulates that

"These shared spaces can become the locus of rich and satisfying experiences in collaborative learning, an interactive group knowledge process in which learners actively construct knowledge by formulating ideas into words that are shared with and built on through the reactions and responses of others" (p. 4).

In an online collaborative learning environment, effective strategies must be laid down to ensure students are not passive but they actively enter into the online classroom and post their thoughts and ideas to the online discussion (Palloff and Pratt, 2004). Moreover, constructivist theory of learning can be supported in Open and Distance e-Learning (ODeL) through a variety of technologies which support constructivist learning like Computer-Mediated Communication (CMC), Computer Supported Collaborative Work (CSCW), case-based learning environments and computer-based cognitive tools (Jonassen et al., 1995). However, social interaction experienced in an online learning environment lacks the face-to-face interaction experienced in a classroom environment (Anderson et al., 2001). Further, there are notable differences between face to face and online interaction like communication limitations due to lack of interaction support tools in real time, and absence of *challenge and explain cycles* of interaction that characterize face-to-face tutorials. This is a limitation to online learning even as its demand continues to rise. To that end, there is a need to carry out more empirical research to maximize the benefits of online learning as in face to face collaborative learning.

In Computer Supported Collaborative Work (CSCW), the communication media also plays an important role as it provides a platform for the group members to discuss adequately and express their ideas in the desirable form. Text based communication like chats, emails, blogs and forums are widely used to discuss group task in an online collaborative learning environment as they are widely available in most LMSs (Muuro et al., 2014b). However, without other visual communication tools such as video link, learners may miss facial expressions which are useful to monitor the partner's understanding of the concept in an online collaborative learning environment. Therefore, if learners are joining groups in different locations, adding a video link may help them understand some cognitive aspects which require some visual expression. This study does not utilize video communication due to both high network bandwidth network requirements and hardware requirements which cannot be afforded by most HLIs in Kenya.

2.6 Analyzing Collaboration in Online Learning

Unlike co-operation which requires "divide and conquer" style of working, collaboration style of working is more complex since it requires interactions among groups to accomplish the same task. However online collaboration can be easier to manage, track and understand because communications can be recorded during the online sessions. Collaboration can be characterized by three important elements: participation, interactions and synthesis (Dillenbourg, 1999). The three elements work together for effective online learning. For example, collaboration requires active participation among its participants such that group members contribute almost equally to the task. Interactions which are collaborative require communications among all group members such that, independent statements and personal comments are avoided (Mason, 1992). The final product must give an indication of group synthesis of ideas and inputs from all members of the group. This section examines some literature on methods for analyzing collaboration with regard to online communications. Also some methods used to measure the characteristics of collaborative learning are discussed.

Online communication for collaboration can take two forms: asynchronous or synchronous (Dillenbourg, 1999). In synchronous communication, there is a time constraint since all participants are required to be online at the same time. Therefore, synchronous communication can only be used in programs like chat rooms, instant messengers, video conferencing programs and other real time message exchanger programs. This limits the use of synchronous discussions to analyze online communication as it is hard to keep track of the constantly changing discussions in a real time discussion. Asynchronous communication occurs without time constraints and therefore, learners have the freedom to discuss at their own free time and from any location, without regard to what other participants are doing. In this study, we focus on asynchronous threaded discussion which is more appropriate to online collaboration because it gives learners a chance to digest the problem and discuss possible solutions for the task (Kaye, 1992). Learning accompanied with online discussions which are asynchronous gives room for learners to extend their classroom learning with deeper discussions of ideas at their own convenient time (Smith, 1994).

Most researchers have used statistics like number of log ons (Ingram, 2000), web server logs (Mason, 1992), number of contributions by both the individual and the group (Barros and Verdejo, 2000), and number of messages (Anaya and Boticario, 2010) to analyze online collaboration. However, literature indicate that little attention has been paid to the analysis of the actual message content due to its complexity (Anaya and Boticario, 2011a). Statistical analysis of messages can be used to give the total number and length of all messages sent and received by all members of the group which can be a measure of individual or group participation in a collaboration task. Ingram and Hathorn (2004) used threaded web discussions in form of asynchronous discussions to analyze online collaboration and it gave positive result for a focused discussion in which a concept is being developed.

Interdependencies (Johnson et al., 1998), synthesis of information (Kaye, 1992) and independence (Laffey et al., 1998) are three critical attributes for a collaborative group. Interdependence is a key element in determining the success of achieving a common goal in a collaborative task (Kaye, 1992). In a positive interdependence, individuals play the role of group promotion to accomplish a group goal rather than individual goal, as the latter goal can only be realized if the first goal is met (Johnson et al., 1998; Kaye, 1992). Synthesis of information in a collaborative group requires members of the group to synthesize the group discussion and give new ideas geared towards achieving a group goal rather than individual goal (Kaye, 1992; Henri, 1992). The independence attribute requires the group discussion to be independent from the instructor where members of the group have the overall freedom to develop a solution for the collaborative task (Ingram and Hathorn, 2004). This can be difficult for learners who are instructor dependent during problem solving (Laffey et al., 1998). If ways are defined to measure the relative amount of the three attributes, then the degree of collaboration among groups can be determined (Ingram and Hathorn, 2004).

Interdependence requires active participation by each member; participation can be measured by counting the number of messages and statements submitted by each individual and the group to the other participants (new posts and replies). This can allow both groups and individuals to be compared in their level of participation. To a lesser extent, ML techniques can be used to compare the current state of interaction with desired state and providence can also be indicated through content analysis of discussion by extracting positive and negative comments about the problem (forum rating by the instructor). Independence on the other hand can be analyzed by measuring the extent of influence by the instructor or other participants in individual participation and interaction (new posts and replies). Individuals who participate and interact without instructor's influence are more independent, hence more collaborative. Synthesis can be measured in two ways: first by the interaction pattern of the discussion that occurs when a participant contributes a statement, another participant synthesizes it by extending the idea and subsequent messages yield new ideas. This requires content analysis of the individual thread contributed in the discussion (forum rating). Secondly, synthesis can be analyzed by examining the final product (learning experience and scores).

Statistical analysis of: (i) dates and times the participants logged on and off, (ii) the order in which the messages are posted, (iii) and the thread in which the messages are placed were used by Ingram and Hathorn (2004) to analyze online discussions. They used these data to diagram the threads of the messages, thus revealing the structure of the conversation. This analysis could only form the basis of a more detailed content analysis of the discussion since analyzing the thread of messages as per the mentioned three criteria was inaccurate and misleading and could not give an accurate indication of collaboration (Ingram and Hathorn, 2004). Although the software they used allowed each student to create discussion threads or to place messages in any specific thread, their results showed poor utilization of the software by the participants.

To get an accurate indication of collaboration, content analysis is important as it can give the extent of collaboration in a web discussion. Research on content analysis for text based communication has been conducted (Silverman, 1993), but some analysis schemes used may be inappropriate for online collaboration (Ingram and Hathorn, 2004) because the online discussion may not have a regular pattern like the written text but evolves with a style determined by the interaction group (Ingram and Hathorn, 2004). Coding schemes designed to measure interactivity do not necessarily measure collaboration (Qing, 2002) unless they measure specific aspects of the discussion using less subjective categories such as problem solving techniques in a specific task, the perception and satisfaction of the participants in the collaboration process (Jonassen and Kwon, 2001). Additionally, since the coding scheme does not provide a measurement model which can assess the discussion directly but rather degree of communication or interactivity, it won't measure the quality of collaboration in the discussion (Gallini and Barron, 2002). Research has shown that if face to face collaboration coding schemes are applied to online collaboration they don't yield good results (Bunnett and Dunne, 1991; Jonassen and Kwon, 2001; Hawkes and Romiszowski, 2001).

Measuring the quality of participation is more relevant in determining the level of participation than quantity (Hiltz, 1990). Therefore, the analyses of transcripts in asynchronous text based communication can be achieved by breaking transcripts into critical thinking units and classifying them into categories which are measurable (Hiltz, 1990). Gunawardena et al. (1997) developed an interaction analysis model which classified messages into one of the five categories:

1. Sharing/Comparing Knowledge
2. Discover/Explore disagreements
3. Synthesis via negotiation meaning
4. Testing/modifying proposed synthesis vs. schemas, theory, facts, beliefs
5. Proofs of reaching agreements or meta-cognitive thus, admitting change of knowledge

Devine (2002) used this model as the basis for developing a set of instructions which clearly identify the expected quality in the discussions to guide participants in discussions forums. This analysis is closely related to Bloom's taxonomy of learning (Bloom, 1956) which suggested six categories of evaluating learner in a learning processing: (i) Knowledge (ii) Comprehension (iii) Application (iv) Analysis (v) Synthesis (vi) Evaluation. The Bloom's taxonomy model can be used to measure the depth of discussions in terms of critical thinking, ability to synthesize other participant's responses and participants' understanding of concepts by automatically analyzing the discussions with tools such as Tag Helper. McLaren et al. (2010) performed content analysis for e-discussions using a Tag Helper to extract specific attributes namely: Unigrams, Bigrams, POS bigrams, Punctuations and Text length which were used to train machine learning algorithms and generate machine learned classifiers of both individual e-discussion contributions and pairs of contributions. The results obtained from this analysis were used to develop a novel Artificial Intelligence (AI)-based graph-matching algorithm. New coding schemes specific to online collaboration discussion are therefore necessary and must be designed to address the variables of interest. Prior determination of categories to be used to evaluate collaboration characteristics, namely; interdependence, synthesis and independence (Ingram and Hathorn, 2004), which enables the researcher to concentrate on other important aspects of online collaboration (Henri, 1992; Miles and Huberman, 1994). For example, Ingram and Hathorn (2004) used the number of messages and statements submitted by each discussant to measure interdependence characteristics, interaction patterns and how the final product relates to the individual group members contributions to measure the synthesis characteristics. Finally, for the independence characteristic, they analyzed the extent to which instructor influences both participation and interaction. Thus, in their research each statement was coded three times to represent three characteristics: interaction, participation and patterns of discussions.

2.7 Group Formation in Collaborative Learning

Group formation is the process of identifying students and assigning them to a specific group so that they belong to one group when doing a group task (Wessner and Pfister, 2001). Assigning students to group membership can be done in a number of ways as summarized in Table 2.2. Groups can either be homogenous or heterogeneous. In homogeneous group formation a student joins a group with other members who have similar characteristics such as course interests, work schedules and residential proximity. For instance, grouping students with interests in the same academic major or with similar course interests may be an effective procedure for promoting bonding, productivity, and synergy among group members, while grouping students with similar class and work schedules can facilitate out-of-class collaboration among teammates. Also, grouping students with respect to residential proximity may be an effective strategy for enabling group members to get together conveniently outside of class to complete group tasks. On the other hand, in heterogeneous group formation a student joins a group with other members who have different or diverse characteristics such as academic achievement, learning styles, personality profiles and demographic information which could include: age, gender, racial and ethnic or cultural background.

Heterogeneous groups are always preferred because it's believed they produce constructive controversy (de Faria et al., 2006). However, though heterogeneous groups are preferred, there is always a dilemma as to what extent there should be heterogeneity in terms of academic achievement, gender, age, social characteristic and personality. Consequently, numerous studies have been conducted to establish the effect of the group formation method on group performance. However, two methods (random selection and self-selection) tend to dominate in the literature, probably due to the fact that there is little involvement of the instructor. However, of these two methods, researchers have posited that self-selection offers the best advantages for students in classroom work groups (Connerley and Mael, 2001; Koppenhaver and Shrader, 2003). The criteria for selecting members in a group can also effect the members' commitment. Group members who choose whom to work with are more relationally satisfied with their group and more committed to work together than members who are randomly assigned to work with each other (Scott, 2001).

Table 2.2: Summary of group formation techniques

Selection Method	Source
By simply determining the group size, then doing random assignment either by the system or random selection by the instructor	(Chapman et al., 2006; Mahenthiran and Rouse, 2000; Muller, 1989)
Self-selection where students make their own choice whom to work with without any direction, interference, or guidance by the instructor	(Scott, 2001)
Instructor selecting group members based on student skills in order to optimize skill distribution among the team. For example, grouping students based on their academic performance.	(Blowers, 2003; Michaelsen and Black, 1994; Zurita et al., 2005)
Assignment based on learners learning style	(Liu et al., 2009)
Assignment based on learners Contextual information	(Messeguer et al., 2010)
Group formation based on learner's profile and learner Context	(Muehlenbrock, 2006)

The random selection method is highly utilized by instructors due to the ease of implementation and 'fair' distribution, which gives a student equal chance to be a member of any group, hence both social and academic heterogeneity can somehow be achieved. However, it can also lead to lack of diversity in skills within the group (Bacon et al., 2001). Randomly selected groups have also proven to utilize their time during group meetings more effectively and group members are more task oriented probably because, familiarity among members is less which makes the groups' social network less compared with self-selected membership (Chapman et al., 2006). Despite these advantages, random selection has proved to be less effective in improving group performance, leads to inferior group dynamic ratings, and results to higher degree of conflicts (Chapman et al., 2006).

Self-selection methods have been reported to improve students' performance in group work than randomly assigned groups (Mahenthiran and Rouse, 2000). Furthermore, this method allows students to: communicate better, have positive attitude towards group work and feel more excited to work together, feel more comfortable to consult one another in their group for help, take more pride in their work and are able to resolve conflicts more effectively than their counterpart in random selection (Chapman et al., 2006). Even so, research on self-selection method has reported that there is a tendency of group members doing others' work (Chapman et al., 2006) and where some students are left out and forced to join others, they may feel they are not part of the self-selected group's social network, what Bacon et al. (2001) refer to as 'remainder problem'.

The use of intelligent systems to do group formation in online collaborative learning environments has also been reported in recent research (Liu et al., 2009; Messeguer et al., 2010). Although computer based random selection methods have been preferred in large classes, intelligent techniques are better because they do incorporate learner's characteristics like learning style (Liu et al., 2009), learner's profile and context (Muehlenbrock, 2006) and contextual information (Messeguer et al., 2010). They could also change the group allocation. The ability to change the group member composition in real time enables the leveling up of learning results and improvements in the participants' social relationships. Some of the intelligent techniques have applied the use of machine learning techniques like Instance-based Learning and Bayesian network which are capable of using contextual information to learn the user behavior and predict an appropriate group for the learner based on the contextual information. Messeguer et al. (2010) and Liu et al. (2009) developed an intelligent grouping algorithm based on learning style and integrated in a LMS to group students with different learning styles together. They also demonstrated the use of it in a realistic online collaborative learning environment by comparing the algorithm with group assignment based on similar learning style. However, in their study they failed to address the impact of the algorithm when compared with other methods such as random and self-selection which are popular in LMS. In addition, there are no true experimental studies on these intelligent systems in order to prove their effect in group performance when compared with instructor based methods.

The place and time which students choose to join a group can also vary depending on the collaborative learning environment. According to Johansen (1988), time space matrix online collaborative learning tools can be classified in four ways as summarized in Table 2.3. Students who are in the same place at the same can join a group and decide to discuss face to face, while those who join a group at the same time but they are in different places can collaborate online using synchronous communication tools such as Chatting, Skype or Video Conferencing. On the other hand if students join a group at different time in the same place or different time and different places, they can collaborate online using asynchronous communication tools such as discussion forums, SMS, Wikis, blocks, etc. This study has utilized asynchronous mode in both cases where learners can meet online at different times at the same place or at different time in different places. This asynchronous mode provides the learners with more time to synthesize the discussion concepts and respond later.

Table 2.3: Time space matrix for collaborative work

	Same time	Different Time
Same place	Face to face interaction	Asynchronous interaction
Different Place	Synchronous Distributed interaction	Asynchronous Distributed interaction

Group size is also another determinant factor in group performance and studies have been conducted to establish the effect it has in group efficiency and outcome. However, there is no consensus on the actual number of group members but most studies have reported that groups with a small size of about 2 to 4 students tend to perform better because of a better sense of responsibility, deeper knowledge of group members and better group co-ordination (Liu et al., 2009). When Students discuss in small groups, they are more satisfied in the learning process and most likely they do show higher academic achievement than those in larger groups (Schellenberg, 1959). But the optimal group size still remains a dilemma as groups with less than four students could also imply that the learning characteristics are not well represented. Some researchers have argued that the larger the group, the greater is the pool of talent and experience available for solving problems or sharing the effort while on the other hand as the size increases, there is higher probability that fewer members will have a chance to participate and probably dominate (Jaques and Salmon, 2007). Therefore, more research and evidence is needed to determine the appropriate group size in learning environments.

Previous research in group member assignment indicates that the method used to assign a member to a group does affect the group dynamics, group performance, group efficiency and effectiveness, and attitude towards group experience (Chapman et al., 2006). Thus, it is critical to evaluate group selection method in relation to all these factors, which most of the current research fails to address. Self selection and random assignment appear to dominate, and hence, the focus of this study. Additionally, majority of researchers in this field fail to use true experimental design approach to evaluate the effectiveness of group selection methods as recommended in educational studies (Fraenkel et al., 2012). This research uses true experiment design approach when comparing the intelligent grouping mechanism with the instructor based methods in order to address the aforementioned shortfall.

2.8 Perceived Challenges in Online Collaborative Learning Environments

Previously, research has been carried out to investigate the learners' satisfaction (Singh, 2005), perceived usefulness and challenges (Song et al., 2004; Kim et al., 2005) and factors leading to unsuccessful group collaboration (Roberts and McInnerney, 2007; Liu et al., 2010), in a collaborative online learning environment. However, results have shown that perceived challenges are likely to vary depending on type of e-learning technology used, infrastructure availability (internet and computers) and the use of different LMSs in HLIs. Furthermore, in Kenya, there is no empirical evidence to establish the perceived challenges in an online collaborative learning environment.

The Kim et al. (2005) study on an MBA online course reveals that even when students had positive attitudes towards online learning because of its benefits (flexibility, more learning experience through social interaction and enhancement of virtual teaming skills) they were faced with some challenges such as difficulty in communication with peers, lack of sense of community and absence of real-time feedback. Existence of these challenges is an indication that learners in this course could not realize the benefits of collaborative learning. In their study, Roberts and McInnerney (2007) identified seven common problems in an online collaborative learning environment: student antipathy towards group work, selection of the groups, lack of essential group-work skills, free-rider, possible inequalities of students abilities, withdrawal of group members and assessment of individuals within the groups. Zorko (2009) investigated factors which inhibit collaboration in wikis and the study provided recommendations on how to increase collaborative behaviors in the wiki in problem based English language learning. Studies have also shown that online learners get frustrated with collaborative learning due to commitment imbalance on the task and lack of common learning goals among students hence requiring the instructor to equip online learners with social and group skills necessary for effective collaboration (Capdeferro and Romero, 2012). Table 2.4 summarizes some of these perceived challenges within three categories: poor motivation, lack of individual accountability and negative interdependence (Liu et al., 2010).

Table 2.4: Summary of perceived challenges in online collaborative learning environments

Category	Description	Source
Poor Motivation	• Posting irrelevant posts to the learning scenario • Misunderstanding the topic • Posts containing grammatical/spelling errors • Difficulty in communication with peers • Absence of real-time feedback • Disagreement among members • Withdrawal of group members • Assignments of students to group membership • Student antipathy towards group work • Lack of common learning goals among students	Liu et al., 2010; Hassanien, 2007; Black, 2005; Capdeferro et al., 2012)
Lack of Individual Accountability	• Not contributing much • Lack of time • Too lazy to work and not meeting deadlines (Free-rider) • Lack of individuals assessment within the groups	(Kim et al. 2005; Liu et al., 2010; Singh, 2005)
Negative Interdependence	• Lack of essential group-work skills • Lack of sense of community • Possible inequalities of student abilities • Single student dominating the group scenario • Unwillingness to critique • Little feedback on each other's work • Commitment imbalance on the task • Poor group management	Liu et al., 2010; Roberts et al., 2007; Capdeferro et al., 2012; Zorko, 2009)

Although most of these challenges are common across the studies, there could be diversity in some cases due to infrastructure availability (like network access, computer-mediated communication tools and instructors support) and student background in different HLIs (Muuro et al., 2014b).

2.9 The Practice of Blended Learning in Kenyan Universities

With the increased demand for higher education in Kenya, e-learning in Kenya has gained popularity. For example, to address the increased demand for e-learning programs in Kenya, recently Kenyatta University (KU) launched a digital school. According to KU website, the digital school offers over 100 courses through blended learning. The students taking these courses can access notes and assignments on the e-learning portal and later they attend four hour face-to-face tutorials for every course before they sit for the final exam. Subsequently, other universities in Kenya have adopted similar learning strategies and now have e-learning portals for blended learning.

With the recent installation of fiber optic cables in Kenya, the cost for internet access and connection has dropped. For example in Nairobi, one can access fiber optic speed of about 100mbps at US $12 per month. According to Kashorda and Waema (2014), about 52% of the students in Kenyan universities own smartphones while 53% own laptops. This shows increased ownership, which coupled with decreased internet access cost means that universities have a good opportunity to offer distance education as well as blended e-learning through technology enhanced pedagogies. The most recent e-readiness survey which was carried out in 17 Kenyan universities indicated that student population doubled within a period of five years, as shown in Table 2.5. Therefore, universities should increase their internet bandwidth expenditures from the current 0.5% to 1.5 % of their total annual expenditure by the year 2016 (Kashorda and Waema, 2014). This was a good recommendation in terms of network access. However, for distance learners to benefit from this bandwidth, pedagogy challenges in e-learning must also be addressed with concrete data within the Kenyan context. This research is timely since Kenyan universities are moving towards digital learning and integration of the Information and Communication Technology (ICT) strategy for universities in Kenya as defined by Kashorda and Waema (2014). Further, there is an increasing demand for online education in Kenya. Therefore, technology enhanced teaching and learning is no longer an option but a requirement to meet the demand for online education. Consequently, the government of Kenya has recommended the establishment of National Open University of Kenya by December 2014, in an effort to expand enrollment through distance and e-learning. Also, there is need to explore other elements in e-learning like collaborative learning which has pedagogical advantages such as development of critical thinking skills, co-creation of knowledge and meaning, reflection and transformative learning (Palloff and Pratt, 2005).

Table 2.5: Demographic data and internet availability for 17 universities in Kenya from 2008-2013

Year of survey	Total students	Total PCs owned by students	Total bandwidth (Mb/s)	Bandwidth per 1,000 students	PCs per 100 students	% of students with PC access at home
2008	162,319	8,907	70.8	0.436	5.5	27
2013	339,418	13,815	1,431.5	4.22	4.07	30.4

Source: KENET e-readiness data in 2008 and 2013

Some universities in Kenya have embraced the use of technology in teaching and they have established institutes like Open, Distance and e-Learning (ODeL) which co-ordinate distance learning programmes, develop e-content and build capacity in e-learning through training staff on e-learning pedagogies and computer centers where distance learners can access online learning materials. The government of Kenya has also established policies to guide ODeL in HLIs which recommends the establishment of an open university as contained in Sessional Paper No. 1 of 2005 (GoK, 2005). This is further indication of the government initiative to support ODeL programs to meet the increased educational demand in HLIs in Kenya. Despite this, previous research in two Kenyan universities (University of Nairobi (UoN) and KU) has identified some key challenges in delivery of ODeL like lack of e-learning resources, higher level of student dissatisfaction (90.8%) and lecturers dissatisfaction (85.6%) with programme organization and delivery (Nyerere et al., 2012). Since these two universities are pioneers in ODeL, these challenges could also be hindering effective implementation of ODeL programs in other HLIs in Kenya.

Although there are many e-learning platforms in Kenya, the most popular ones are Moodle and Blackboard which do provide both synchronous and asynchronous collaborative tools. Using these e-learning platforms, learners are able to follow lectures online, interact with lecturers, start online discussions through various collaborative tools, submit assignments and check on their academic progress online. Despite the potential benefits of collaborative learning like: development of critical thinking skills, co-creation of knowledge and meaning, reflection and transformative learning, these collaborative tools are yet to be put into full utilization as according to Nyerere et al. (2012), with most of the instructors using the e-learning platforms to communicate to their students. Some private universities in Kenya such as Strathmore University and United State International University (USIU) have adopted the use of e-learning in more than 80% of their courses, while public universities such as KU has only managed to offer about 25% of their course through Moodle platform. This information is found on the universities' websites. Evidently, private universities are utilizing e-learning platforms more than public universities (Muuro et al., 2014b).

2.10 Introduction to Machine Learning

Machine Learning (ML) is a field in computer science which deals with computer programs which learn from experience how to perform task and are expected to improve their performance with time. Mitchell (1997) defines machine learning as follows:

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E, pg. 2.

Therefore a machine learns whenever it responds to the environment in such a manner that it's expected future performance improves (Witten and Eibe, 2000). Therefore ML requires a machine being capable of improving its performance based on its past experience. For example when the performance of speech recognition machine improves after hearing several samples of a person's speeches, then we can justify it has learnt how to do speech recognition. Just like in human beings, after learning has occurred, ML algorithms can be used to make intelligent decisions based on some data. This makes the field of ML learning to fall under artificial intelligence track since they are used to solve problems by emulating the human learning process. Thus, ML algorithms have proven to be of great practical value in a variety of application domains such as: speech recognition, data mining, pattern recognition, predictive models, decision based on uncertainty, classification problems among others. ML techniques have also been preferred in solving certain tasks which are not easy to be solve through human beings. Some of these are:

1. Those tasks which cannot be defined well except by example; that is, we might be able to specify input/output pairs but not a concise relationship between inputs and desired outputs. In this kind of situation ML techniques are capable of adjusting internal structure to produce correct outputs for large number of sample inputs and thus suitably constrain the input/output function to approximately the relationship implicit in the examples. This makes ML useful in areas such as pattern recognition like face recognition from images, speech recognition, associating student behaviors with personality, etc (Mitchell, 1997).

2. In some cases some important relationships are hidden among large piles of data making it difficult for human beings to extract these relationships. This concept in ML has been referred as data mining (Witten et al., 2011) and it has a number of applications in educational set ups like mining education data in collaborative learning to discovering learning patterns and diagnosing students problems (Anjewierden et al., 2007).

3. When the amount of data is too huge for explicit encoding by humans, ML

techniques become valuable as they are capable of learning from huge databases through different mechanisms like use clustering algorithms, association rules, classifiers, probability models, etc.

4. Some tasks are done within a dynamic environment which requires continuous adaptation to the environment and constantly discovering new knowledge which may prove difficult to human beings. ML techniques do prove to be good in domains where the program must dynamically adapt to changing conditions like learning interests (Mitchell, 1997). This makes it possible to integrate ML techniques in e-learning platforms and provide personalized/adaptive learning (Lu et al., 2007; Gong, 2008; Li and Chang, 2005; Surjono, 2014; Esichaikul et al., 2011) or discover student preferences (Carmona et al., 2007).

Due to this wide application, ML becomes a multi-disciplinary field which draws on ideas from a diverse set of disciplines such as: artificial intelligence, computational complexity, information theory, neurobiology and psychology, control theory, probability and statistics and philosophy. ML algorithms are organized into a taxonomy, based on the desired outcome of the algorithm. These taxonomies share similar characteristics with human learning. Some of these taxonomies which are commonly applied in learning include:

1. Supervised learning: in this type of learning the task is to find a deterministic function that maps any inputs to desired outputs such that disagreement with future input-output observations is minimized. For example, in a classification problem, the learner approximates a function mapping a vector into classes by looking at input-output examples of the function. Clearly, whenever asked for the target value of an object present in the training sample, it is possible to return the value that appeared the highest number of times together with this object in the training sample. However, generalizing to new objects not present in the training sample is difficult. Example problems are classification and regression. Example of relevant algorithms include; Logistic Regression, Decision trees, Naive Bayes

classifier, Bayesian Network, Support Vector Machine and the Back Propagation Neural Network.

2. Unsupervised Learning: It deals with unlabeled data where the learning algorithm is given a training sample of objects, for example images or pixels, with the aim of extracting some "structure" from them – e.g. identifying indoor or outdoor images, or differentiating between face and background pixels. One of the most general ways to represent data is to specify a similarity between any pairs of objects. If two objects share much structure, it should be possible to reproduce the data from the same "prototype". This idea underlies clustering algorithms: Given a fixed number of clusters, we aim to find a grouping of the objects such that similar objects belong to the same cluster. We view all objects within one cluster as being similar to each other. Example problems are association rule learning, evolutionary learning paradigms and clustering. Example algorithms but not limited include: Apriori algorithm, Self-Organizing Maps (SOM), Expectation Maximization (EM) and k-means.

3. Reinforcement Learning: The problem of reinforcement learning is to learn what to do and how to map situations to actions so as to maximize a given reward. In contrast to the supervised learning task, the learning algorithm is not told which actions to take in a given situation. Instead, the learner is assumed to gain information about the actions taken by some reward not necessarily arriving immediately after the action is taken. One example of such problem is learning to play chess. Each board configuration i.e. the position of all figures on the 8×8 board is a given state; the actions are the possible moves in a given position. The reward for a given action (chess move) is winning the game, losing it or achieving a draw. Example algorithms include: Q-learning and temporal difference learning.

2.11 Clustering Algorithms

Clustering is the process of finding out a group of objects which have similar characteristics and assigning them to a cluster such that objects in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis. The principle of clustering is maximizing the similarity among the object groups in a cluster and minimizing the similarity between the object groups in different clusters (Romero et al., 2008). Clustering methods can be classified into different types (Jain et al., 1999), including hierarchical (single-link, complete-link, etc.) and objective-function-based algorithms (K-means, expectation maximization, etc.). These clustering algorithms are available in Weka software which is open source software implemented in Java code and platform independent. In Weka the algorithms can be applied directly to a dataset or invoked through other software (WEKA, 2007). The Weka workbench also provides a graphical interface which allows easy visualization of data and other explorers for managing data.

In e-learning, clustering has a number of applications including but not limited to the following:

1. Finding clusters of students with similar learning characteristics and to promote group-based collaborative learning as well as to provide incremental learner diagnosis (Tang and McCalla, 2005) .

2. Defining clusters of students based on performance in certain activity e.g. active and non active students (Aher and Lobo, 2011).

3. Discovering learning patterns which reflect user behaviors in order to characterize similar behavior groups in unstructured collaboration spaces (Talavera and Gaudioso, 2004).

4. Grouping students according to their collaboration competence level in a

collaborative learning environment (L. Valetts and Gesa, 2008).

5. Grouping students in order to give them differentiated guidance according to their learning skills and other characteristics (Hamalainen et al., 2004).

6. Grouping tests and questions into related groups based on the data in the score matrix(Spacco et al., 2006).

7. Predictions of student's academic performance.

For the purpose of this study, in the following section we explore K-means and Expectation Maximization.

2.11.1 K-Means Clustering Algorithm

K-means clustering is an algorithm used to classify data based on attributes into k number of clusters, where k is a positive integer. The grouping is simply done by minimizing the sum of squares of distances between data and the corresponding cluster centroids. Therefore, given a data set of n data points $x_1, x_2, \ldots, x_n$ such that each data point is in $\mathbb{R}^d$, the problem of finding the minimum variance clustering of the dataset into k clusters is that of finding k points(cluster centroid) $\{m_j\}(j = 1, 2, \ldots, k)$ in $\mathbb{R}^d$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\left[min_j d\left(x_i, m_j\right)\right] \tag{2.1}$$

is minimized, where $d\left(x_i, m_j\right)$donates the Euclidean distance between x_i and m_j. The points $\{m_j\}(j = 1, 2, \ldots, k)$are known as cluster centroids.. Therefore, in Equation 2.1 the problem is to find k cluster centroid, such that the average squared Euclidean distance(mean squared error, MSE) between a data point and its nearest centroid is minimized. The k-means algorithm becomes a gradient descent procedure, which begins at starting cluster centroids, and iteratively updates these centroids to decrease the objective function described in Equation 2.1.

In summary, the k-means algorithm can be described as follows:

1. Enter the number of clusters (value of k) to group data as *k* points into the space represented by the objects that are being clustered.
2. Calculate the arithmetic means of each cluster formed in the dataset and assign each object to the group that has the closest centroid
3. When all objects have been assigned, recalculate the positions of the *k* centroids
4. Repeat step 2 and 3 until convergence is achieved , that is until the centroids no longer move.

k-means when compared with other clustering algorithms has some advantages such as:

1. It's simple to implement
2. It's computationally faster when dealing with large number of variables than hierarchical clustering provided the value of *k* is small
3. It may produce tighter clusters than hierarchical clustering, especially if the clusters are globular

However, it do possess the following disadvantages:

1. Different values of *k* may affect the outcome, making it difficult to compare the quality of clusters
2. When the number of clusters is fixed it may be difficult to predict the *k* value
3. It does not work well with non-globular clusters

2.11.2 Expectation Maximization (EM) Algorithm

EM algorithm is an iterative method for finding maximum likelihood estimates of data distribution when data is partially missing or hidden. Suppose we have data elements (x_1.........x_n) grouped into three clusters c_1, c_2 and c_3. The process of grouping these data elements can be summarized into two steps:

1. Expectation (E) step: Which estimates the probability of each element belong to each cluster i.e. $p\left(c_j|x_k\right)$. Each element is composed by an attribute vector(x_k). The relevance degree of the points of each cluster is given by the likelihood of each element attribute in comparison with the attributes of the other elements of cluster c_j.

$$p\left(c_j|x\right) = \frac{|\sum_j(t)|^{-\frac{1}{2}} \exp^{nj} p_j(t)}{\sum_{k=1}^{m} |\sum_j(t)|^{-\frac{1}{2}} \exp^{nj} p_k(t)} \tag{2.2}$$

The EM iteration alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood evaluated using the current estimate for the parameters as shown in Equation 2.2, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the (E) as shown in Equation 2.3. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

Where, x is input dataset

$$P(t+1) = \frac{1}{M}\sum_{k=1}^{M} P\left(C_j \mid x_k\right) \tag{2.3}$$

M is the total number of clusters, t is an instance and initial instance is zero.

2.12 Application of Machine Learning in Collaborative Learning

The use of the new tools for online communication by instructors and different instructional design techniques has greatly improved the online collaborative learning (Bonk and Dennen, 1999; Freeman, 1997). Although instructors encourage students to form group discussions not all of them will lead to successful collaboration unless, the instructor monitors the students' interactions and communicates with the aim of improving the collaboration process. For online collaboration, the instructor may not be able to monitor the collaboration process due to time limit and the huge number of geographically dispersed students. This necessitates the use of intelligent systems which can analyze and modify the online collaborative process dynamically on behalf of the instructor (Israel and Aiken, 2007; Liu et al., 2009).

Recent researches on the use of ML techniques which can emulate the role of an instructor to improve the collaboration process have been conducted with a positive outcome. Anaya and Boticario (2010) and McLaren et al. (2010) in their research project (ARGUNAUT), they used two ML algorithms: j48 decision trees and ada boost with decision stamp to analyze past e-discussions. Results from the analysis of past e-discussions were used to provide "awareness indicators" for instructors in the context of new e-discussions. In their analysis they applied two shape coding mechanisms; (i) shape-level coding scheme which primarily focused on interpretation of the text within a shape and (ii) paired-shapes coding schemes which involved analysis of structural, process-oriented, and textual aspects of the shapes. They define seven annotation variables (topic focus, task-management focus, critical reasoning, request for clarification or information, critical evaluation of opinions, summary and intertextuality) related to shape level and five annotation variables (question answer, contribution followed by question, contribution-counterargument, contribution-supporting argument, and qualify/comprise) related to paired-shapes level.

Among the attributes they used, critical reasoning yielded good results. Their ultimate goal was to train machine learning classifiers to predict the appearance of these discussion characteristics in the context of a new discussion and use the resultant classifier to provide awareness indicators in their project. To validate the machine learning techniques they used two freely available softwares: TagHelper and YALE. With 10-fold cross validation and with no parameter tuning, Adaboost with decision stump gave the best performance. However, to achieve high generalization for the classifiers, they recommended the use of more extensive machine learning experiments with bigger groups which can provide larger corpus for training the ML techniques.

Messeguer et al. (2010) on their research for group prediction in collaborative learning applied two machine learning algorithms: (i) Instance Based Learning (IBL) based on the fact that it is simple algorithm which tolerates noise, can cope with irrelevant attributes and can exploit inter-attribute relationship; and (ii) Bayesian Network (BayesNet) based on the fact that it's a typical and well known learning algorithm. They used logon information to capture features like: (i) timestamp in terms of the time and day of week, (ii) user identify in terms of username and Bluetooth MAC address, (iii) place based on the access point, and (iv) neighborhood obtained from a list of Bluetooth MAC address, but they did not address the use of individual learner's cognitive characteristics like knowledge level to estimate group membership. Data collected from these features was transformed to input vectors and output vectors to create a training model for the two machine learning algorithms. In their training model, IBL gave 100% accuracy while BayesNet gave 85% accuracy. They also applied cross-validation technique to validate their algorithms before testing with new dataset. With new testing data set IBL gave 95% accuracy while BayesNet gave 70% accuracy, an indication that IBL has better generalization on group prediction task.

Anaya and Boticario (2009b) applied clustering algorithm EM to reveal relation between the statistical indicators which are related to learner's collaboration (number of messages sent/replied and threads in a conversation initiated by the learner) and the learner collaboration. These statistical indicators were domain-independent. They used the clustering algorithm EM to group the learners according to their level of collaboration based on a dataset created with instances from the mentioned statistical indicators obtained from student's interactions in discussion forums. However, there was no attempt to compare the performance of the clustering algorithm EM with other ML algorithms like K-NN or decisions tress. To evaluate the performance and accuracy of the clustering algorithm EM, clusters created by the algorithm were compared with an expert list obtained form expert analysis on the same task. From their results, clustering algorithm EM performed almost similar to the expert analysis, but they recommended more research on (i) methods which can provide sufficient information on the collaboration process and (ii) carrying a parallel research with other ML algorithms like decision trees.

Anaya and Boticario (2011a) applied two machine learning algorithms: (i) clustering based on the fact that with sufficient data it can group instances with no prior knowledge of the relevant attributes (Gama and Gaber, 2007) and (ii) decision trees based on the fact that it is well known algorithm capable of learning a given classification, to analyze student interactions and improve the collaboration process. Clustering algorithm was used to group students according to their collaboration (high, low or medium) while the decision tree algorithm was used to assign a collaborative value to each student and compare them. They used twelve statistical indicators which were based on the number of threads or conversations initiated in team's forums and the number of messages sent/replied by the student. The later statistical indicator was found more relevant. A performance evaluation for the accuracy of the two algorithms similar to their earlier research (Anaya and Boticario, 2009b) was performed. Although they compared the two algorithms, an empirical analysis with others was not performed and their application in group formation has not yet been tested.

2.13 Conceptual Framework

The conceptual framework is defined in terms of definition of conceptual elements, relationship between independent, intervening and dependent variables and operationalization of the variables.

2.13.1 Definitions of Conceptual Elements

Independent Variables

The independent variables in this study are derived from group formation techniques. Three different group formation techniques are studied, which include: random assignment, Grade Point Average (GPA) and intelligent grouping. These three different group formation techniques are used to construct our independent variables. In random assignment, group members are assigned at random and therefore, random numbers are used as indicators. In GPA method, students' performance in a given period of time is used as an indicator. In intelligent grouping, collaboration competence level is used as an indicator whereby data mined from a discussion forum is used to cluster students based on their collaboration competence level.

Dependent Variables

Our dependent variables are derived from the group outcomes. The group outcomes include the group performance, learning experiences, perceived group problems, group task satisfaction and group leader satisfaction. These five different group outcomes are used to construct our dependent variables. Performance in group work can be characterized by three characteristics, namely: interdependence, synthesis and independence. Indicators for these group outcomes include: number of new posts/

replies in a discussion forum, forum rating scores assigned by an instructor and scores obtained from a written test/quiz related to the discussion forum.

Figure 2.1 illustrates the relationship between independent and dependent variables.

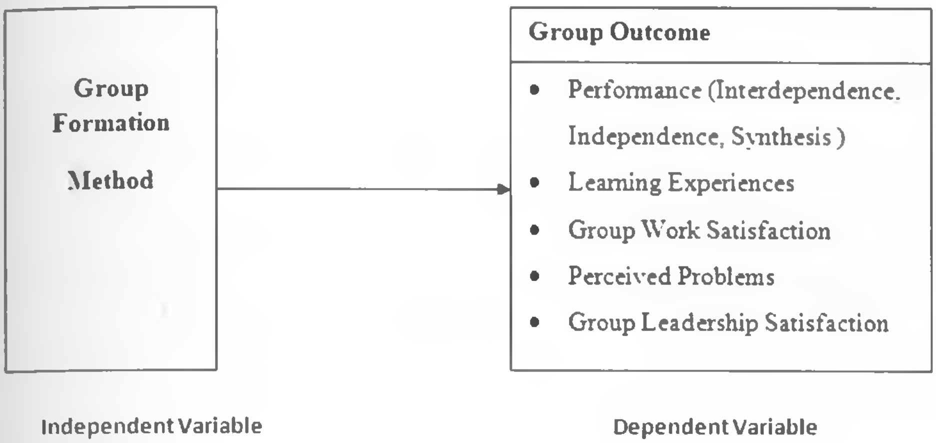

Figure 2.1: Conceptual Framework

2.13.2 Operationalization of Variables

In order to measure collaboration competence level we introduce three collaboration characteristics, namely: interdependence, independence and synthesis. Interdependence requires active participation by each member; participation can be measured by counting the number of messages and statements submitted by each individual and the group to the other participants. This allows both groups and individuals to be compared in their level of participation. Independence, on the other hand, can be analyzed by measuring the extent of influence by the instructor or other participants in individual participation and interaction. Individuals who post new ideas rather than just replies are more independent hence, more collaborative. Synthesis can be measured in two ways: first by the interaction pattern of the discussion that occurs when a participant contributes a statement, another participant synthesizes it by extending the idea and subsequent messages yields new ideas. This requires content analysis of the individual thread contributed in the discussion forum. Secondly, synthesis can be analyzed by examining the relationship between original comments and the final product. In this study, we apply the latter where the instructor compares the post with the final product and assigns a numerical value as per the relevance. This in turn can tell us the level of individual contribution in relation to the final product. A summary of the three critical attributes of collaboration, categories used to evaluate them and their measurement criteria is provided in Table 2.6.

Table 2.6: Characteristics of collaboration, their related parameters and the measurement criteria

Characteristic of collaboration	**Category used to evaluate it**	**Parameters/ indicators to be measured**	**Measurement criteria**
Interdependence	Participation	Messages and statements	Counting the number of individual messages and statements in the discussion
	Interaction	Comments in the discussions	Counting negative and positive comments
		Statements/ contributions	Length of statements/depth of the contribution
		Individual contributions	Percentage of contributions responded or linked to other contributions by the individual, other than the contributor.
Synthesis	Interaction and final product	Initiative, creativity and conformity	Arguments, agreements/ disagreements and proposals on contributions like new ideas.
Independence	Participation and interaction	Patterns of discussions	Counting direct responses, direct comments, and indirect comments
		Statements	Independence of the statement (its connection to previous statements e.g. agreement or disagreement)

In the light of the above arguments, we apply the three attributes to define three collaboration competence levels (High, Medium, Low) which are characterized by different levels of interdependence, synthesis and independence as described in Table 2.7.

Table 2.7: Characteristics associated with collaboration competence levels

Collaboration competence level	**Characteristics**
High	If a student logs in often, participate and interact actively and indicates high level of interdependence, synthesis and dependence. His profile is clearly collaborative and the learner can be ranked into a higher level of collaboration competence.
Medium	If a student logs in often , participate and interact moderately and indicates moderate interdependence, synthesis and dependence. His profile is medium and the learner can be ranked into a medium level of collaboration competence. At this level the learner needs assistance to improve to high level.
Low	If a student logs in and participates rarely and there is no indication of interdependence, synthesis and dependence. His profile is non collaborative and the learner can be ranked into a low level of collaboration competence. At this level, the learner needs immediate attention to improve to medium level.

Operationalization of variables which are indicated in the conceptual framework is shown in Table 2.8

Table 2.8: Operationalization of variables in terms of Indicators, Measurements Criteria and Scale

Variable Type	Main Variable	Sub Variables	Indicators/ Measuring criteria	Values	Scale
Independent	Group Formation Method		Student performance (GPA)	Class one	Nominal
			Random Numbers	Class three	
			Collaboration Competence Level	Class two	
Dependent	Group Outcome	Performance • Interdependence • Independence • Synthesis	• Quiz • Written test • Forum rating	Scores	Ratio
		Learning Experiences	• Number of post/replies • Ratings	5- point likert scale	Ordinal and Interval
		Group Work Satisfaction	• Number of post/replies • Ratings	5- point likert scale	Ordinal and Interval
		Perceived Problems	• Number of post/replies	List of group problems	Ordinal
		Group Leadership Satisfaction	• Number of post/replies • Ratings	5- point likert scale	Ordinal and Interval

2.14 Summary

Literature has shown that collaborative learning finds its roots from social construction of knowledge. Thus, collaboration becomes an important aspect of learning for the constructivist pedagogy. In collaborative learning, the goal is to create social interaction which will result to the acquisition or construction of new knowledge. According to Vygotsky (1978), collaboration is more fruitful when learners collaborate with experts or more able peers because what an individual does jointly with others can be incorporated into his individual problem solving process. Based on Vygotsky (1978) ZPD concept, Tudge (1990) explained the effectiveness of collaboration by stating that: *children who were led to think at a higher level through being paired with a more capable peer achieved that higher level in the course of collaboration and generally retained it in subsequent independent performance* (p.163). Engaging students in collaborative learning has been recognized as a powerful method to motivate learners.

Similarly, collaborative pedagogy in the digital technology has changed the ways in which students interact with their instructors within the learning process. As noted in the literature review, research has been conducted to address collaborative pedagogy in the digital technology (Gunawardena et al., 1997; Harasim et al., 1995; Barrows, 1994; Koschmann, 1996; Jonassen, 1997; Palloff and Pratt, 1999). With the realization of digital technology, online collaborative learning provides new opportunities for student collaboration in an online environment and new challenges for teachers supporting group work Bonk et al. (1998); Palloff and Pratt (1999).

The digital technology has provided online tools like discussions forums, chat rooms, e-mails, newsgroups, etc., which provide a collaborative learning environment. For instance, from the literature it has been noted that: (1) discussion forums create a platform where learners can learn on their own with the opportunity of sharing experiences and construct knowledge based on their cognitive level (Corich and Hunt, 2004), (2) with e-discussion forums it's possible to have social affective and cognitive benefits of face to face situations realized (Hiltz, 1990), (3) if discussion forums are managed well it becomes a major tool for supporting e-learning as they encourage learners to share knowledge and build new ideas from shared concepts (Garrison, 1993). Online tools for group activities like the discussion forums and chat rooms allow learners to build self-esteem, learn to accommodate diverse opinions on issues, enhance their listening and communication skills and develop skills needed in team workforce (Johnson, 1984; Taylor, 2004).

CHAPTER THREE

METHODOLOGY - SYSTEM DEVELOPMENT

3.1 Introduction

In developing this system, a multi-methodological approach was used. This system development methodology consists of four research strategies: theory building, experimentation, observation and system development as illustrated in Figure 3.1.

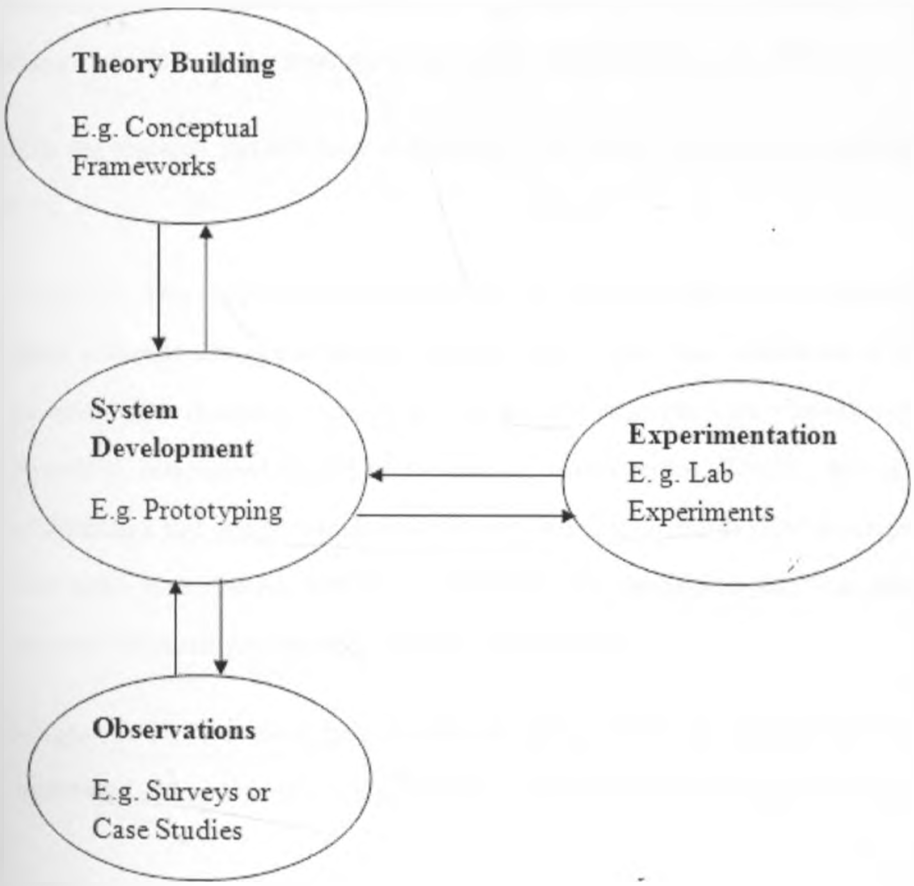

Figure 3.1: Multi-methodological approach to system development

In this methodology, system development is viewed as the hub of research that interacts with other research methodologies to form an integrated and dynamic research process (Nunamaker et al., 1991).

In theory building developments of new ideas as contained in conceptual framework are required to guide the system development process. Observations which are done through research methods such as surveys do provide desirable information in the research domain which can guide the system development process and the experimentation process. Experimentation provides a chance to validate the system and the underlying theories. These methodologies complement one another hence, providing valuable feedback to one another. In case of complex research areas such as intelligent systems, multi-methodological approach becomes an effective strategy for gaining a complete understanding of the system (Nunamaker et al., 1991).

Therefore, the research methodology in this study cuts across two inter-related chapters as follows:

1. Chapter 3: This chapter first, describes how the system development requirements were gathered through a survey whereby a pre-study was conducted in order to inform the design process of the system and also the experimental design. Secondly, conceptual design of the system is discussed. Thirdly, the system architecture and design which describe how the intelligent module is integrated into LMS is presented. Finally, it describes how the prototyping was done to develop the intelligent module and how it was tested.

2. Chapter 4: This discusses the experimental design which was applied in this study to evaluate the system in a realistic online collaborative learning environment.

3.2 Pre-Study

This study was conducted to investigate the current status of online collaborative learning in Higher Learning Institutions (HLIs) in Nairobi, Kenya, and identify perceived challenges in an online collaborative learning environment. Nairobi County was selected because e-learning infrastructure is more established in terms of network access due to fiber optic network and education demand are higher as compared with other regions in Kenya. Two primary questions guided this pre-study:

1. What is the current status in terms of collaboration tools, group orientation and collaboration activities on online collaborative learning in HLIs in Nairobi, Kenya?
2. What are the components of online collaborative learning which learners perceive as challenging in HLIs in Nairobi, Kenya?

A cross-sectional survey was used to investigate the current status of online collaborative learning in terms of collaboration tools, group orientation and collaboration activities and the related challenges in HLIs in Nairobi, Kenya. A descriptive survey was adopted as it could examine the situation the way it is and provide quantitative information that was summarized through statistical analysis, thus providing the basis to answer our research questions (Engelhart, 1972). The researcher administered questionnaires using a web-based tool (Limesurvey). This approach was preferred because it enabled a faster collection of responses and the ease of exporting the data to the Statistical Package for Social Sciences (SPSS) for analysis.

3.2.1 Target Population

Purposive sampling was adopted to select two public universities: Kenyatta University (KU) and Jomo Kenyatta University of Science and Technology (JKUAT), and two private universities: United State International University (USIU) and Australia Studies Institute (AUSI), which have adopted the use of online collaborative learning tools in their e-learning modules and they are within Nairobi. To identify our target population, instructors who were teaching online and they had engaged their students in group activities online were requested to provide emails of these students. With the help of these instructors, a total of two hundred and ten students were identified within the four universities. These students were enrolled in at least one course or a module online on an e-learning platform. These students were informed by their instructors of the purpose of the study, and responded to the questionnaire items voluntarily.

3.2.2 Data Collection Instruments

Data was collected through a questionnaire that consisted of thirty items. The literature review provided the conceptual elements which were used to develop the set of items in the questionnaire. Twenty nine items in the questionnaire were close ended while one item was open ended. Table 3.1 summarizes the different categories for the questionnaire items. To ensure validity, content related evidence was used and two experts in e-learning were requested to review the content and the format of the instrument. Based on their comments, some of items were rephrased, some content in group orientation added and reformatting done as recommended. Content-related evidence was adopted to ensure the instrument contained adequate sample of the key challenges related to online collaborative learning (Fraenkel et al., 2012).

Table 3.1: Summary of the questionnaire items

Item Number	Type	Information Gathered
Items 1-7	Multiple choice	Demographic information
Item 8	Multiple choice	Gadgets used by students to access online materials
Item 9	Likert Scale	How often a collaborative tools is used to do collaborative work
Items 10-11	Multiple choice	To filter students who had participated in an online group activity so that they could proceed with item 12 up to 30
Item 12	Multiple choice	Frequency of use on the collaborative tools
Items 13-21	Multiple choice	Group orientation in terms of how the groups were formed, managed and students' satisfaction with their group membership
Item 22	Multiple choice	Instructor's role during the group activity
Items 23, 24, & 25	Multiple choice	Level of individual participation in the group activity
Items 22, 26, & 28	Multiple choice	Student experiences during the group activity
Items 27 & 29	Likert Scale	Student level of agreement on group work challenges as observed from literature review.
Item 30	Open ended	Students' worst experiences in an online collaborative group activity from their own perspective

3.2.3 Data Collection and Analysis

The questionnaire was distributed through email invitations to the participants. The invitation email contained the purpose of the study and a link to the URL where the questionnaire was located. Each participant was given only one token to ensure a single response to the questionnaire. The questionnaire was made available for a period of two weeks as most of the students did not respond immediately. A total of 183 students responded: - this was an 87% response rate which was adequate for analysis. The collected data was coded and exported to SPSS for statistical analysis.

3.2.4 Survey Results

3.2.4.1 Participants' Demographic Information

A total of 183 students responded out of 210, with 44.9% from a private university and the 53.5% from public university while three respondents (1.6%) did not provide university names. One respondent did not provide any demographic information including age and gender. Table 3.2 summarizes the demographic information.

Table 3.2: Demographic information of the sample (n=183)

Characteristic	Frequency	Percentage
1. Age in bracket(at the time of survey)		
15-25 years	108	59.00%
26-35 years	59	32.20%
36-45 years	13	7.10%
46-55 years	2	1.10%
N/A	1	0.50%
2. Gender		
Male	116	63.40%
Female	66	36.10%
N/A	1	0.50%
3. University		
AUSI (Private)	14	7.70%
JKUAT (Public)	50	27.30%
KU (Public)	48	26.20%
USIU (Private)	68	37.20%
No Answer	3	1.60%
4. Level of Study		
Certificate	14	7.70%
Diploma	5	2.70%
Postgraduate	21	11.50%
Undergraduate	142	77.60%
No Answer	1	0.50%
5. Modules Studied online		
2-3 modules	35	19.10%
4-5 modules	27	14.80%
More than five modules	51	27.90%
One module	68	37.20%
No Answer	2	1.10%
6. Internet Skills		
Excellent	138	75.40%
Good	32	17.50%
Moderate	11	6.00%
No Answer	2	1.10%

3.2.4.2 Group Characteristics

Data was collected on five group characteristics which included: (i) criteria used to assign group membership, (ii) number of members in the group, (iii) whether there was any change in group membership during the entire course, (iv) involvement of a moderator or mentor and (v) how comfortable a member was in the group. Table 3.3

summarizes the results on these characteristics. The findings indicated that assignment to group membership was done in different ways and also group sizes were also different. Out of 108 students who responded, a higher percentage of group assignment was done by the instructor (59%), 16% at random through default assignment in Moodle, 18% self assignment and 7% were not aware how the assignment was done. The number of students in a group ranged from 2 to 5 (32%), 6 to 10 (27%) and more than 10 students (35%). While 6% were not aware of the number of students in their group. This shows that more than 50% of students discussed in groups of more than five students which is contrary to the recommended small group sizes of 2 to 5 students for effective group learning which enables each group member to express his own ideas and increases group cohesion (North et al., 2000; Schellenberg, 1959; Forsyth, 2009). Furthermore, 81% of the students remained in the same group during the entire period of the course. Only 58% had the instructor as the moderator/mentor while 18% had the student playing the role of the moderator/mentor in their discussion forums. The rest 24% had no one to play the role of a moderator/mentor. Notably, 93% reported they were comfortable with their group membership, but three participants (3%) were not comfortable and they had varied reasons which included:

1. Respondent one: *"lack of familiarity made it easy to lie to one another"*
2. Respondent two: *"they didn't pull their weight"*
3. Respondent three: *"people I didn't know then took advantage of my hard work"*

Table 3.3: Summary on group characteristics

Group characteristics (N=108)	Frequency	Percentage
1.Criteria used to assign group membership		
Assigned by Instructor	64	59%
Default assignment in Moodle	17	16%
I assigned myself	19	18%
I don't know	8	7%
2.Number of members in a group		
2- 5 members	34	32%
6-10 members	29	27%
More than 10 members	38	35%
I don't know	7	6%
3.Change in group membership during the study of the unit/module		
Yes	27	25%
No	81	75%
4.Moderator or the Mentor person in the group activity (N=82)		
Instructor	63	58%
Student	19	18%
I don't know	26	24%
5.Membership Comfort ability within the group (N=103)		
Yes	100	93%
No	3	3%
I don't know	5	4%

3.2.4.3 Popularity of Various Collaborative Tools

As shown in Table 3.4, of all the respondents, 91.8%, 74.8%, 72.9% and 71.9% frequently use email, social media (Facebook and Twitter), telephone (mainly mobile phones), and chats respectively. Tools like Skype, Video conference, Workshops [1] and Podcasts [2] had the lowest frequency of use, which is an indication that these tools are

[1] Peer assessment activity in Moodle
[2] Audio files created by students for peer learning

rarely used by students to collaborate online. Table 3.4 shows the percentage, mean ranking and standard deviation on the frequency of use on various collaboration tools.

Table 3.4: Frequency of use on various collaborative tools

Collaboration tool	n	Rarely	Often	Mean	Std. Deviation
Emails	182	8.20%	91.80%	0.92	0.276
Social Media	182	25.20%	74.80%	0.75	0.436
Telephone	181	27.10%	72.90%	0.73	0.446
Chats	181	28.10%	71.90%	0.72	0.451
Google Doc.	180	47.20%	52.00%	0.53	0.501
Wikis	178	65.80%	34.20%	0.34	0.476
Forums	181	67.40%	32.60%	0.33	0.47
Skype	182	72.50%	27.50%	0.27	0.448
Video Conference	181	84.00%	16.00%	0.16	0.368
Workshops	178	84.30%	15.70%	0.16	0.365
Podcasts	178	93.80%	6.20%	0.06	0.241

3.2.4.4 Level of Collaboration in Various Collaborative Tasks

Out of 183 students who responded, only 108 students (59%) indicated that they had done some group work online in their e-learning modules. The rest of the respondents (41%) were not involved in an online group work for reasons which included: Instructor not providing an online group activity (41.3%), lack of time (29.3%), lack of skills to participate in online discussion (12%) and not enrolling to a group (17.3%).

More than 80% of the respondents had very low access to posts and they were not replying to posts; only less than 20% accessed or replied to posts more than 4 times in a week. It was found that: 39.8% of the respondents indicated that either they accessed or replied to posts only once in a week, 42.7% accessed the posts 2-3 times in a week, 48.5% replied to posts 2-3 times in a week. Table 3.5 summarizes the observed level of access and reply to posts.

Table 3.5: Students level of access and reply to posts in an online group activity (n=108)

No. of times of accessing and sending posts to the discussion forum	Access to posts		Sending new posts/replies	
	Frequency	**Percent**	**Frequency**	**Percent**
Only Once	41	39.80%	41	39.80%
2-3 times in a week	44	42.70%	50	48.50%
4-5 times in a week	6	5.80%	7	6.80%
More than five times in a week	12	11.80%	5	4.90%

3.2.4.5 Perceived Challenges

The questionnaire item on the perceived challenges had nine challenges which respondents' were required to rate with a yes or no response. The study revealed that the majority of respondents (54%) perceived that lack of participation by other members was a big challenge as most students lacked time to participate (53%). The difference in skills or knowledge level among group members was not perceived as big challenge (19%). Table 3.6 shows the distribution of responses on the nine key challenges from 108 respondents. In addition to these nine key challenges, slow internet connectivity (30%), disruptions from incompetent peers (3%), lack of clarity on the posted work (2%), free-riders (2%), no consensus on the discussions (3%) and no original ideas posted (5%) were also mentioned by respondents as some of their worst experiences during their group work. For example, participant number 9 stated: *"My worst experience was when the internet was not consistent and it kept logging users ON and OFF; and we ended up wasting almost one hour without active participation"*.

To establish whether there was any relationship between the type of university (public or private) and the perceived challenge, chi-square test of independence was done. Table 3.7 summarizes the results of the chi-square test and the corresponding p-values. Statistical significance of association was found only in two cases; Lack of feedback from instructor (p =.041) and workload not shared equally (p =.000).

Table 3.6: Mean ranking and standard deviation for the nine key challenges (n=108)

Challenges	Mean	Std. Deviation
Low or no participation of other group members	0.54	0.501
Lack of time to participate	0.53	0.502
Lack of feedback from instructor	0.47	0.502
Lack of feedback from peers	0.43	0.497
Off-topic posts in the discussion	0.31	0.463
Work load not shared equally	0.27	0.445
Lack of group mentor	0.25	0.435
Single student dominating	0.25	0.435
Difference in skill/knowledge level among group members	0.19	0.390

Table 3.7: Associations between University type (private or public) and the perceived challenges

Perceived Challenge	x^2	p
Low or no participation of other group members	0.255	0.613
Lack of time to participate	0.4	0.527
Lack of feedback from instructor	4.176	0.041*
Lack of feedback from peers	0.844	0.358
Off-topic posts in the discussion	0	0.99
Work load not shared equally	12.802	0.000*
Lack of group mentor	1.913	0.167
Single student dominating	0.004	0.947
Difference in skill/knowledge level among group members	0.179	0.672

3.2.5 Pre-Study Outcome

Firstly, the study aimed to investigate the status of online collaborative learning in HLIs in Nairobi, Kenya. The findings indicate that out of 183 respondents who were doing module/unit through e-learning, only 108 students (59%) were engaged in an online group activity, while 75 students (41%) were not involved in an online group work. The study found that failure of the instructor to provide an online group activity contributed highly to non-participation in collaborative learning. Moreover, for those who participated in group work, 47% mentioned that they perceived lack of feedback from the instructor as a big challenge. There was an indication that online instructors were not fully engaging students in collaborative learning in blended e-learning programs. This could be due to the known situation that in most HLIs in Kenya, instructors use e-learning platforms to send notes and assignments, are heavily burdened with many duties and they lack skills in e-pedagogy (Nyerere et al., 2012). The study also found that for those who were engaged in collaborative work, the level of collaboration was very low as most of the respondents (90%) accessed the discussion forum less than two or three times in a week. Consequently, the rate of posting to the discussion forum was found to be very low with only 11.7% of the respondents sending an average of 4 to 5 posts in a week. The findings also revealed that students in these HLIs do not often collaborate online. Hence, there is a need to hire more trained instructors or train the current instructors who can engage students in an online collaborative work and create time to monitor their participation.

Secondly, the study aimed to investigate the components of online collaborative learning which learners perceive as challenging in HLIs in Nairobi, Kenya. The findings indicated that 54% strongly perceived that lack of participation by other members was a big challenge. This could be supported by the fact that 53% of respondents did not have time to participate. Lack of feedback both from instructor and peers was also perceived as challenge by 47% and 43% respectively. This concurred with results from other researchers who found that low participation by members and lack of feedback both from instructor and peers was a major hindrance to collaborative learning (Liu et al., 2010; Capdeferro and Romero, 2012; Kim et al., 2005). Although Roberts and McInnerney (2007) identified seven common problems, to the contrary in this study the problems were not major as few respondents were in agreement. Furthermore, 30% of the participants mentioned slow internet connectivity as one of their worst experiences even though previous research had not captured it. This could be due to low internet bandwidth (4.22Mb/s per 1,000 students) availed to students in Kenyan Universities (Kashorda and Waema, 2014). This was somewhat surprising given that the study was conducted in Nairobi where internet infrastructure is far much better than other regions in Kenya where fiber optic network is yet to be established. This implies that for other regions the problem will be more critical. Thus, we concur with Kashorda and Waema (2014) proposal in their e-readiness report to have HLIs in Kenya invest more in campus backbone and wireless network infrastructure to increase the level of internet availability to students.

Thirdly, there was a significance different between the public and the private universities in terms of lack of instructor support ($p = 0.046$) and workload not shared equally ($p = 0.000$). The study found that lack of instructor support as challenge was reported more in public universities (31%) than in private universities (16%). This could be due to the big numbers of students enrolled per class in public universities which makes the instructor to student ratio higher than in private universities. Consequently, the instructors in public universities are overloaded with work and this could have affected the low level of feedback observed. The study also found that the challenge of workload not shared equally among the students in an online collaborative learning group was reported to be higher in private universities (20%) than in public universities (7%). This seems to support the perception that students in public universities are more independent, working with minimal instructor supervision, which probably gives them an advantage to work more cohesively in group work.

Table 3.8 summarizes the key findings and how they guided the system development process.

Table 3.8: Summary of key findings and how they guided the system development

key Finding	How it guided the system development process
Moodle is commonly used as an e-learning platform	The LMS to be used was Moodle because of its popularity
Group formation was not effective as majority of the students were assigned groups at random by the instructor or self selected their own groups. Therefore, learner's characteristics were not considered during group formation.	Design an intelligent grouping algorithm based on learners' collaboration competence level which required minimum instructor support/intervention and groups students dynamically with the desired group size. This could accommodate learner's characteristics in group formation
Instructors do not provide adequate online group work. Consequently, in blended learning the full potential for collaborative learning is yet to be realized	Instructional design on the course to be studied to include collaborative learning tasks.
Lack of peer participation and instructor feed back was a major challenge. Consequently, the level of collaboration was very low among participants	SMS and email platform to be added in Moodle to allow immediate feedback to be provided by the instructor
Commonly used online collaborative tools include social media , mobile phones, emails, chats and forums	Forums to be used for online discussion as they do allow asynchronous discussion which most of the students prefer because of time dynamics.
Slow internet connectivity was a major problem hindering effective online collaboration	Actual study was carried out with students who had full access to campus network.

3.3 System Development

In order to develop the system through the multi-methodological approach, a tool building process which consists of five stages as proposed by Nunamaker et al. (1991) was adopted. These stages include:

1. Conceptual Design: This stage involves the development of a conceptual framework. This framework helps the researcher to formulate the important concepts and develop a theory to support the system development.

2. System Architecture: This stage involves the development a system architecture.

The system architecture provides a blue print for the system building process. It also defines the system components in terms of functionality, structural relationships and dynamic interactions among the components.

3. System Design: This stage involves analyzing and designing the system. This involves understanding the studied domain, the application of relevant scientific and technical knowledge, the creation of various alternatives and the synthesis and evaluation of proposed alternative solutions.

4. Prototyping: This stage involves the development of a prototype system in order to test the system in a real world setting. This prototype can further be developed into a final product and implemented if full functionality is realized.

5. Experiment: This stage involves system evaluation through experimentation in real world settings. Through experimentation researchers can observe the functionality, performance and the impacts of the system on individuals, groups or organizations.

Figure 3.2 illustrates the five stages and a description on how the were adopted to fit the system development in this study is provided alongside. This chapter addresses the first four stages in details while stage five is addressed in the next chapter.

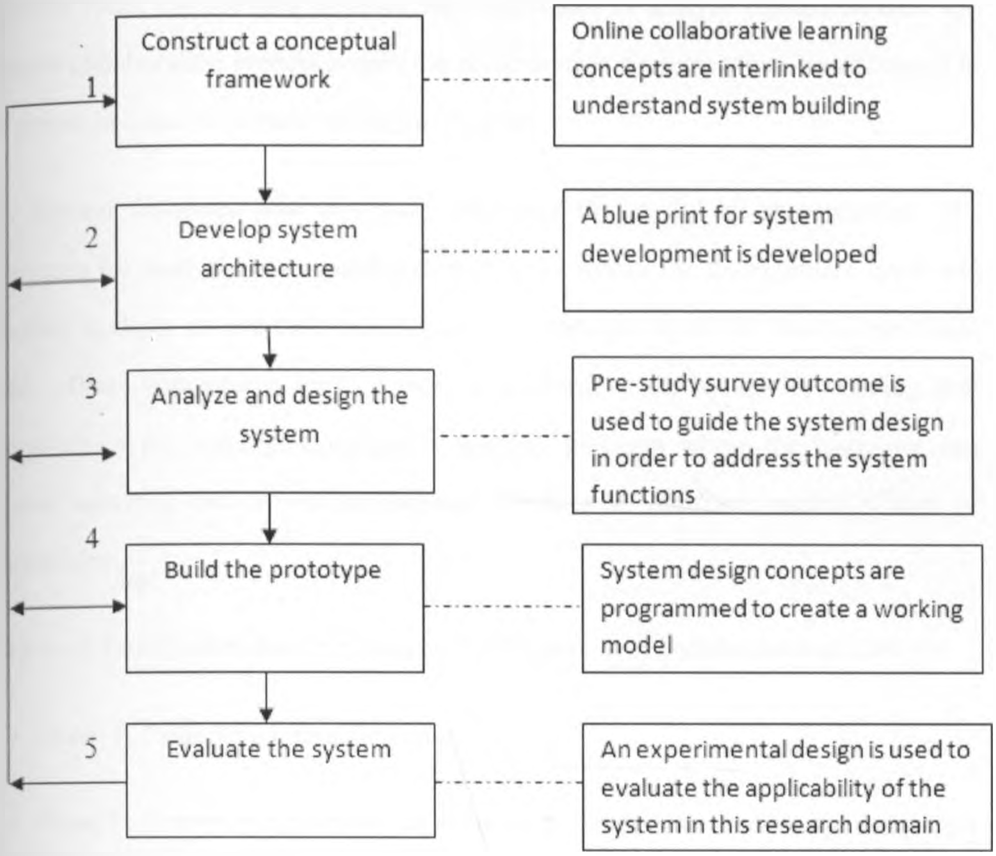

Figure 3.2: System Development Research Methodology Process

3.3.1 Conceptual Design

As noted in the literature review, there is lack of good methodologies and standards to analyze online collaboration (Anaya and Boticario, 2011a). Most of the tools also discussed in previous studies do provide little support to the instructor in managing the collaboration process. A good conceptual model for analyzing quality data and assist in managing collaboration proces should consider both statistical analysis (quantitative) and content analysis (qualitative). In this study, we adopted a conceptual model based on management collaboration cycle framework (Soller et al., 2005). This framework discusses how to support collaboration process by the use of mirroring and meta-cognitive tools, but the idea of using ML techniques to analyze discussion data and improve collaboration process within the collaboration cycle needs to be addressed in this model in order to provide intelligent support.

This section discusses how this study improves the model by incorporating ML techniques for analyzing the collaboration process within the management cycle and classifies students as per their collaborative competence level (L. Valetts and Gesa, 2008). These competence levels provide a platform in the model for creating new groups which are heterogeneous and a learning platform where the instructor can propose activities and provide customized feedback to reinforce student's level of collaboration.

Soller et al. (2005) identifies five phases of management of collaboration process:

- Phase 1: Collecting interaction data
- Phase 2: Constructing a model of interaction: This phase requires the computing of indicators to represent the current state of interaction. Statistical tools can be used to analyze interaction data and provide these indicators
- Phase 3: Compare current state of interaction to desired state
- Phase 4: Advise/guide the interaction, and
- Phase 5: Evaluate interaction through assessment and diagnosis

The interaction of these phases is illustrated in Figure 3.3

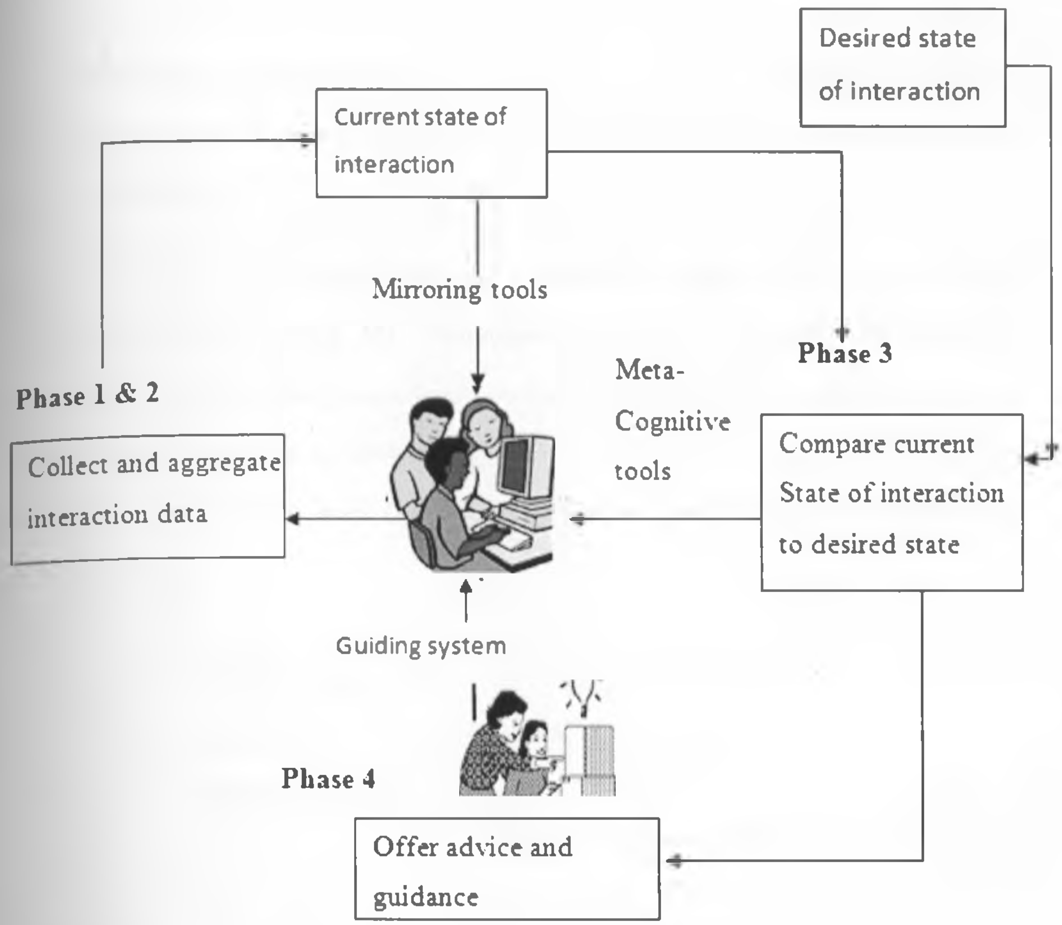

Figure 3.3: A Framework for Collaboration Management
Source: Soller et al. (2005)

This framework explains how to manage collaboration process, but it lacks a guideline on how to incorporate ML techniques into the collaboration management process. Through this framework, ML techniques has been incorporated in this collaboration management cycle with a view of improving the following phases:

- Phase 2: ML techniques can be used to analyze interaction data and cluster data as per the leaner's collaboration competence level (L. Valetts and Gesa, 2008).
- Phase 3: ML techniques can be used to compare the current state of interaction with desired state and reveal indicators which are related to learner's collaboration (Anaya and Boticario, 2009b).

- Phase 4 and 5: Results from ML techniques can be used to define collaboration competence levels. The later can be applied to automate the group formation process and dissemination of feedback to reinforce student's level of collaboration.

In the light of the above arguments, we formulated a framework which analyzed collaboration process using ML techniques and model learner's collaboration competence level, initialize groups based on learner's collaboration competence level and disseminate feedback as shown in Figure 3.4. The next section describes how collaboration competence level can be defined in an online collaborative learning environment.

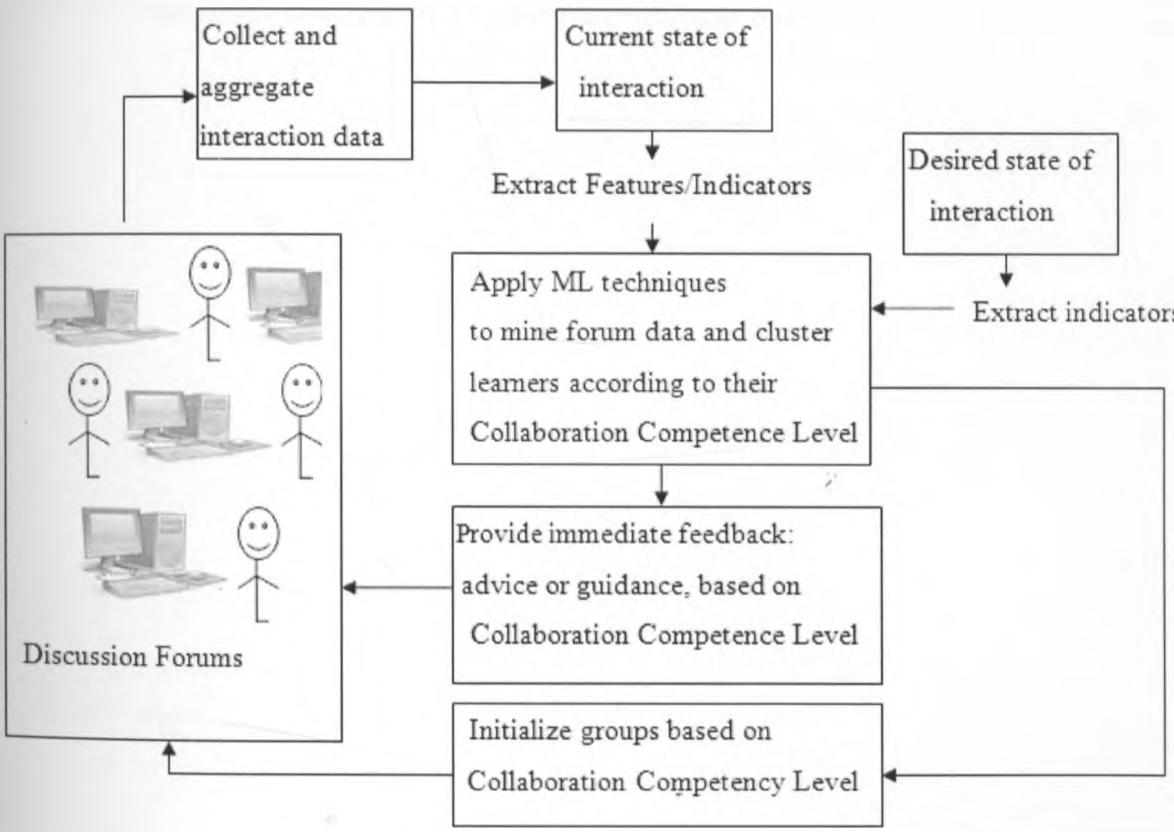

Figure 3.4: Conceptual Model for integrating ML techniques into collaboration management cycle

3.3.2 System Architecture

In this section, we demonstrate a system architecture that integrates ML algorithms into LMSs such as Moodle. In order to use ML to support discussion forums in Moodle, first the system architecture for Moodle is linked to ML environment. The ML environment contains the clustering algorithms which are applied to the preprocessed Moodle forum data obtained from Moodle Database(DB) to create clusters which are equivalent to the number of collaboration competence levels defined by the instructor. The data for the resulting clusters is post-processed and stored back to Moodle DB. These cluster data is applied by the intelligent grouping algorithm to create groups for collaborative work and on the feedback platform to disseminate customized feedback based on learner's collaboration competence level. Figure 3.5 illustrates this system architecture.

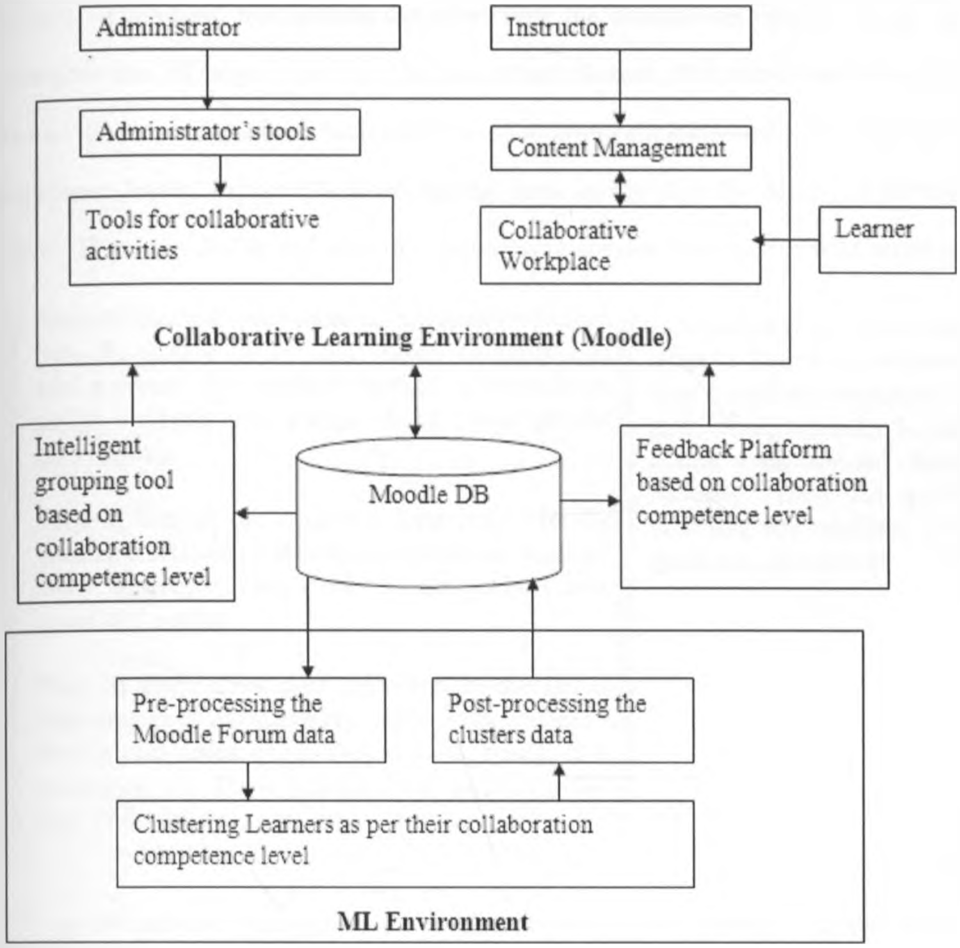

Figure 3.5: System Architecture for ML support to group work

3.3.3 System Design

The design process was guided by the pre-study outcome. From the pre-study, it was confirmed that Moodle e-learning platform was commonly being used and therefore the researcher utilized this platform to design group formation process and feedback based on the learners' collaboration competence levels. The pre-study informed the researcher to focus on group formation and feedback since group formation was not effectively done. Evidently, most students discussed in groups of more than five without getting immediate feedback through emails or SMS. Currently, Moodle can only group students automatically through the random method and it does not provide an SMS platform. Therefore, this section describes how the design was carried out in order to integrate the ML algorithm into Moodle, create clusters of students based on forum data and develop a database which can be used to group students based on collaboration competence levels. Figure 3.6 illustrates the steps involved in the design of the entire system . Database design and user interface design are discussed in the next section.

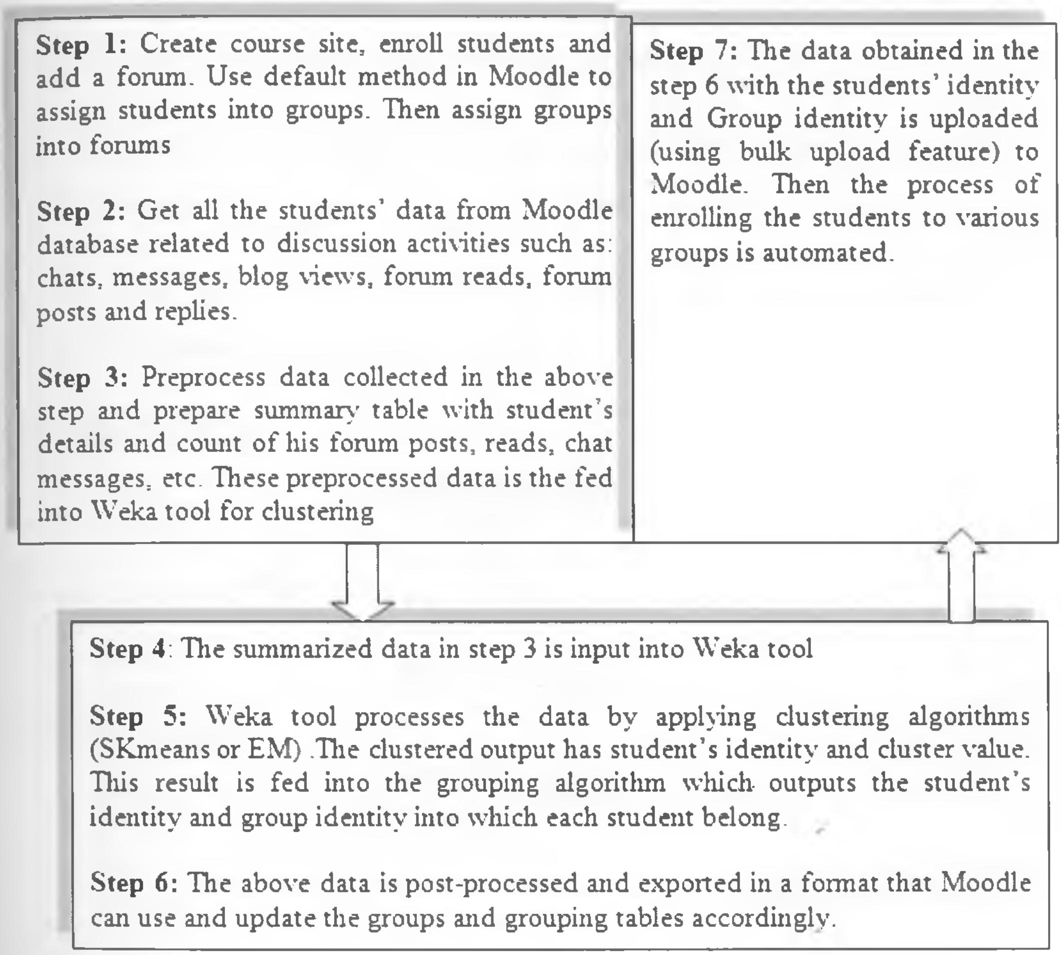

Figure 3.6: Design process for group formation using intelligent grouping algorithm

(a) Database Design

Four important tables are defined in Moodle where two of the tables deal with definitions of groups and groupings and the other two are for taking care of membership, assigning users into groups and groups into groupings. These four tables include: (i) *mdl_groups*, (ii) *mdl_groupings*, (iii) *mdl_groupings_groups and* (iv) *mdl_groups_member*

The relationship for these tables is illustrated in Figure 3.7. In order to create the intelligent grouping algorithm using the cluster's data, the following tables are added in Moodle database:

1. *mdl_cluster_temp*: contains the attributes(*id, userid, clustered*)
2. *mdl_groups_temp* (temporary table with same mdl_groups structure to store temporary groups)
3. *mdl_groups_members_temp* (temporary table with same mdl_groups_members structure to store temporary group members)

Figure 3.7: ERD for group tables in Moodle database

(b) User Interface Design

In order to execute the grouping and the feedback module, a graphical user interface has been added in Moodle. This interface has *'Discussion Groups'* menu which the instructor can use to view, manage clustering and create groups using the intelligent grouping algorithm. The *'Intelligent Grouping based on Learners' Collaboration (IGLC)'* grouping menu when executed loads a form which has an option for creating clusters and groups. The clustering option has been designed with options to select the type of clustering algorithm (SKMeans or EM) and number of clusters. The grouping option has been designed with the clustering options together with the number of groups option. Forms are also designed to display both cluster and group information. For each cluster, text fields are added on the form to display: number of clusters and number of students in each cluster together with the student's identity number (*student_ID*). For the group information the form has text fields to display the group identity number (*group_number ID*), group mentor identity number (*group_mentor_ID)* and group membership information for each group. Forum statistics menu has also been designed to allow the instructor to view an individual's forum statistics in terms of number of new posts, number of replies and the rating scores. This interface design is illustrated in Figure 3.8.

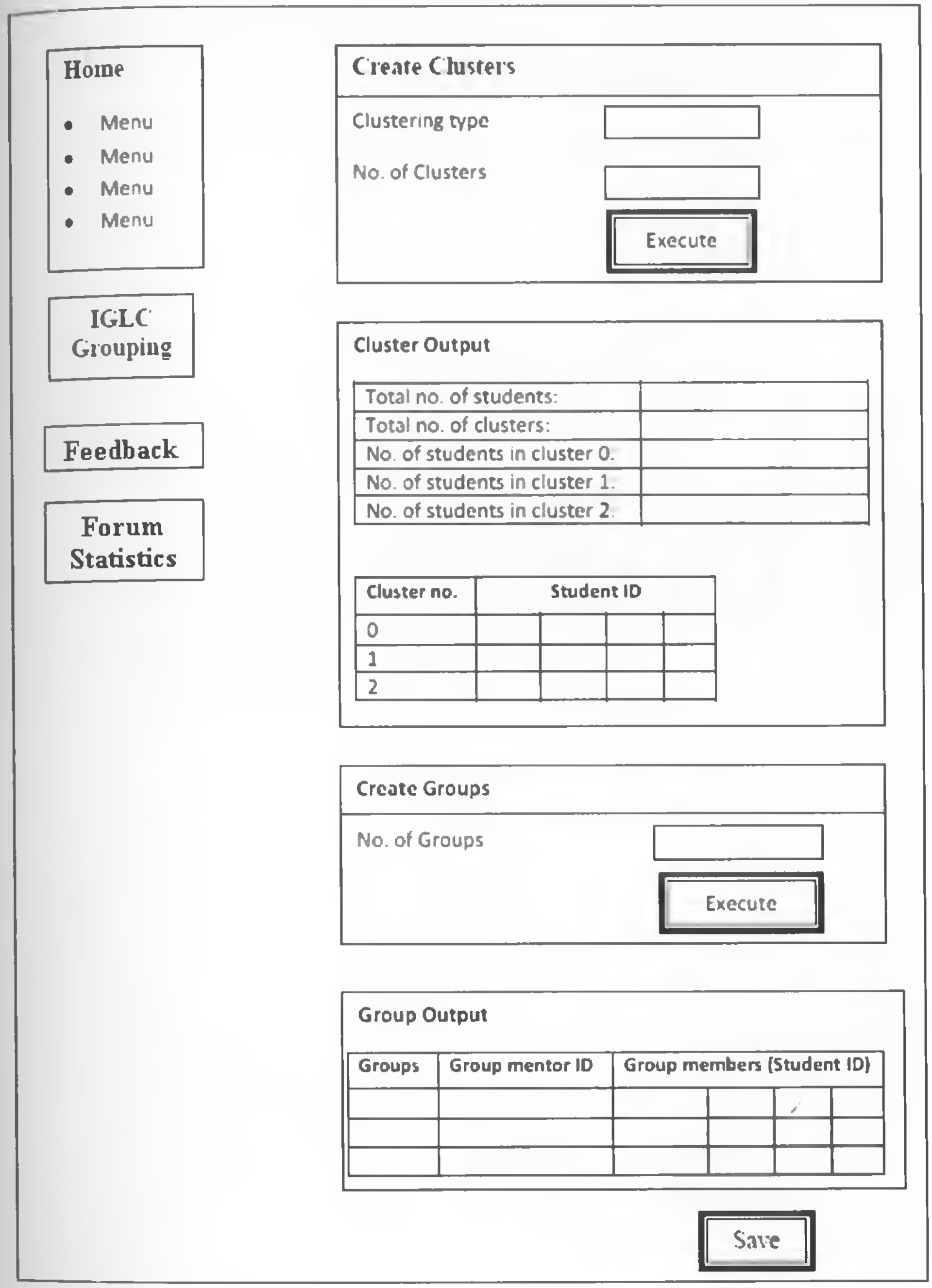

Figure 3.8: User interface design

Forms are also designed to provide a feedback platform which allows the instructor to broadcast the message either to the whole cluster or to individual students and also to select the type of message mode (email or SMS). The Figure 3.9 illustrates this design.

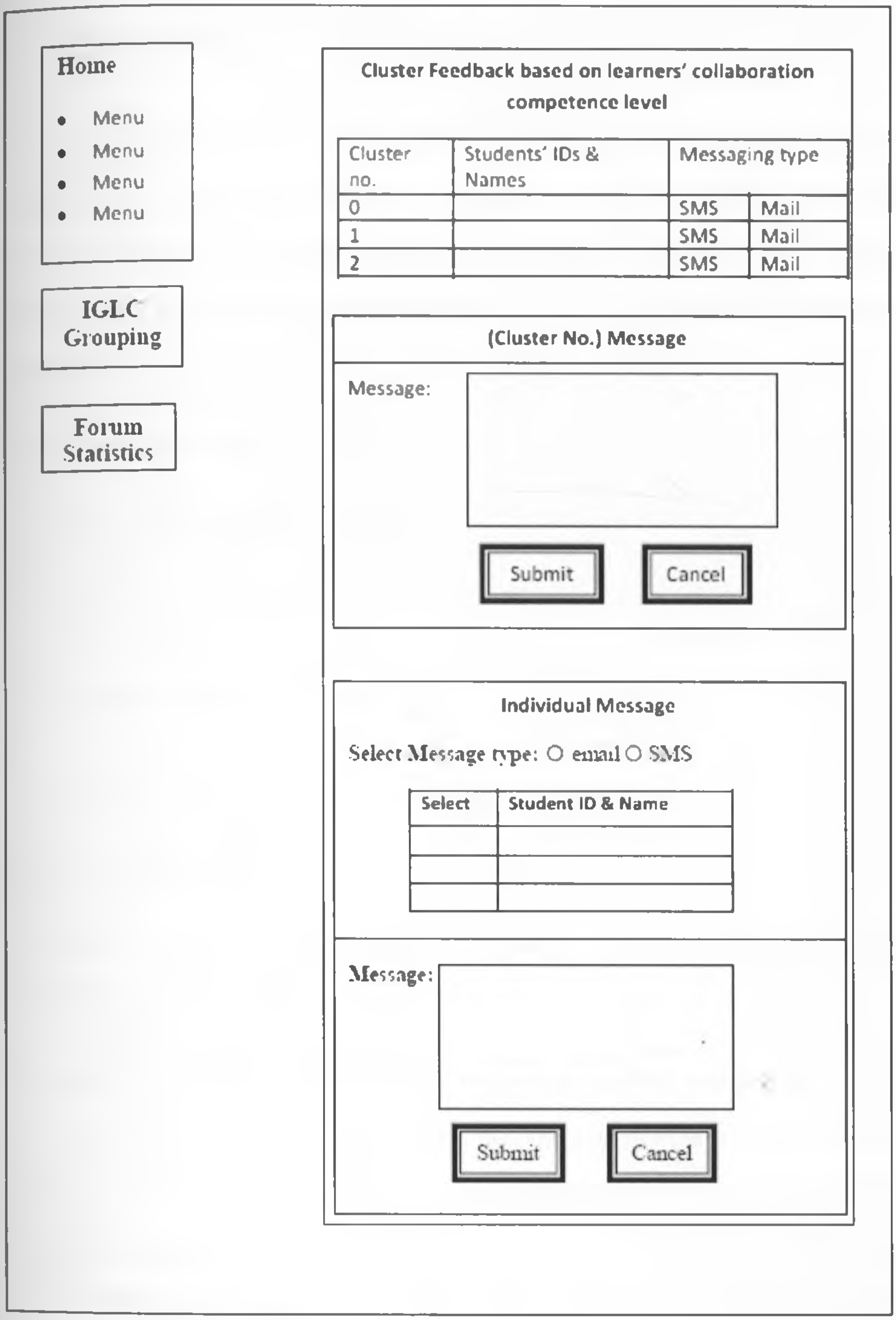

Figure 3.9: Messaging interface design

3.3.4 Prototyping

This involved the actual coding of the system and its integration into Moodle. In order to code the system, firstly the implementation tools were identified and those which were not part of Moodle were integrated. Secondly, the required coding was done as per the design and finally, testing of the system was done. This section therefore, discusses the following areas:

1. Implementation tools
2. Cluster implementation in Moodle
3. Intelligent grouping algorithm
4. Feed back platform (email and SMS)
5. Prototype testing

(a) Implementation Tools

In this study, the following components were utilized in order to develop the prototype and realize the research objectives:

1. Moodle e-learning platform: Moodle e-learning platform was selected because: (i) It is an open-source learning course management system which is utilized by the larger community in higher learning institutions, (ii) the availability of its source code makes it possible to have it customized as per the user requirements, (iii) Moodle is designed with a number of activities such as chat, forum, glossary, wiki and workshop which do support collaborative learning and (iv) stores discussion forums data in a Relational Database Management System(RDBMS) which is easy to manipulate.

2. Forums: These are discussion platforms in Moodle where an instructor or student

can post a discussion on specific topic and students can engage in asynchronous discussion online. Rating of the posts can also be done either by the instructor or the students themselves. Forums were used because they do offer the following advantages:

- Discussion forums create a platform where learners can learn on their own with the opportunity of sharing experiences and construct knowledge based on their cognitive level (Corich and Hunt, 2004).
- With e-discussion forums it is possible to have social, affective and cognitive benefits of face to face situations realized (Hiltz, 1990).
- When a discussion forum is well managed, it becomes a major tool for supporting learning as it encourages learners to share knowledge and build new ideas from shared concepts (Garrison, 1993).
- Online tools for group activities like the discussions forums and chat rooms allow learners to build self-esteem, learn to accommodate diverse opinions on issues, enhance their listening and communication skills and develop skills needed in team workforce (Johnson, 1984; Taylor, 2004).

3. Moodle Database: Moodle stores detailed information for all activities that students perform (Rice, 2011). Therefore data related to forums has been extracted from Moodle database.

4. Machine learning tools: The study used clustering algorithms based on the fact that, with sufficient data they can group instances with no prior knowledge of the relevant attributes.Two clustering algorithms (SKMeans and EM) from Waikato Environment for Knowledge Analysis (Weka) software have been integrated in Moodle to analyze forum data and form clusters based on learner's collaboration competence level.

5. Application Programming Interface (API) for SMS: *'Africastalking'* API has

been integrated in Moodle to provide SMS services.

6. Weka software which has number of ML algorithms for clustering, has been used to provide clustering algorithms utilized in this study. Weka software was chosen because it is an open source software in which its Java code can be invoked within the Moodle PHP code environment and has a set of well organized collection of state-of-the-art ML algorithms and data pre-processing tools.

   To integrate the Weka software into Moodle, the following was done:

   (a) Moodle 2.3 was installed on the Windows Server

   (b) Weka Jar lib was added to the Windows Server

   (c) Weka Jar lib was invoked from the Moodle PHP code

(b) Cluster Implementation

The Weka software has several clustering algorithms available. However, in this study we only used SKMeans and EM clustering algorithms. The objective was to group students into 3 clusters based on discussion forum data in Moodle. Therefore, this section discusses how the forum data is pre-processed and fed into Weka.PHP program. The forum data in Moodle is stored in MySQL Moodle database. Although forums data have many attributes, we have utilized three attributes which possess data that corresponds to the three indicators of collaboration (Interdependence, Independence and Synthesis). The first attribute is a new post which is an original idea; the second is a reply to post which corresponds to a response to an existing idea and the third is average rating of the posts which indicates the level of relevance of the post on the issues under discussion.

Preprocessing the data requires the data to be cleaned and transformed into an appropriate form which can be processed by Weka clustering algorithms. Moodle forum data and forum rating is stored in the following tables: *mdl_forum:* stores information about all forums; *mdl_forum_posts:* stores all posts to the forums; *mdl_forum_discussions:* stores all forums' discussions and *mdl_rating:* stores the average rating of the posts

Since the data is stored in a RDBMS, less cleaning and pre-processing is required and for our case, we only create a summarization table with the required fields from the above tables and export the result to a text file with *'CSV'* format which is applicable to Weka tool. To create the summarization table, the following SQL statement is executed.

```
SELECT role.userid ,count(if(post.parent=0 ,post.userid ,NULL)) AS
    Numberofpost ,
count(if(post.parent!=0 ,post.id ,NULL)) AS Numberofreplies ,
(select round(COALESCE(avg(rate.rating),0))
from mdl_rating AS rate where role.userid=rate.userid and
rate.component='mod_forum' and rate.ratingarea='post') AS avgrating
FROM mdl_context AS context INNER JOIN mdl_role_assignments AS role
ON role.contextid=context.id and role.roleid=5
LEFT JOIN mdl_forum_posts AS post ON role.userid=post.userid
WHERE context.instanceid=$courseid and context.contextlevel=50 group
    by role.userid
```

The summary table is stored as text file with *.cvs* extension and it has the following columns:

1. User id (taken from mdl_role_assignments by checking the role and enroll conditions)

2. Number of posts (taken from mdl_forum_posts)

3. Number of replies (taken from mdl_forum_posts)

4. Forum ratings (taken from mdl_rating)

Dynamically this data is exported into CSV format file and stored in same directory. This summary table is fed as an input to the Weka.PHP program which has the clustering algorithm. The Weka.jar lib is invoked within the Weka.PHP page in Moodle. The Weka program takes following input parameters; Input file, Type of Clustering (Skmeans or EM) and number of clusters (3). The result is 3 clusters along with cluster sizes and *userid* in each cluster. The number of clusters correspond to the collaboration competence levels defined earlier (High, Medium, Low).

(c) Intelligent Grouping Algorithm

Data stored in the three clusters was used to form heterogeneous groups using an intelligent grouping algorithm. To create heterogeneous groups, the data stored in the three collaborative competence levels (cluster 0, cluster 1, cluster 2) is converted to an array with '*userid*' values. A randomizing algorithm created using php '*randomarray*' function takes the array as input and produces an output array with randomized '*userid*' values. For example, if cluster 0 corresponds to higher collaborative level and has '*userid*' values as per this order: 12, 34, 56, 23, 47 then after randomization the order changes to: 34, 47, 23, 56, 12. This randomization task is done for all clusters and then '*userids*' are ranked from cluster 0 (most collaborative) to cluster 2 (least collaborative). The result is stored in an array called '*rankedArray*'. It's from the '*rankedArray*' the algorithm picks students from different collaborative levels as per the rank and assigns them to one group as per the specified group size. The process is performed iteratively until all students are assigned to a group. Students who are most collaborative are assigned a mentor role in their group.

The following pseudocode was applied to implement the intelligent grouping algorithm based on clustered data.

```
start_session := load_csv_file < filename(mdl_cluster_temp)
declare variable and initialize()< inputs int(i,j,n,a,b,
    no_of_cluster,userst,no_of_groups,rank)
```

```
declare variable and initialize()< inputs array(random_array,
   new_array,test_array,group_array)
// store cluster asignment in double dimensional array:(Array[i][j])
foreach(no_of_cluster);
userst=<get_recordset>;
                    foreach(userst)
                              Array[i][j]=userst;
                              j++;
                                                  i++;
// randomize the double dimensional array
for(i=0;i<n;i++)
random_array=Array[i];
// randomize the array by using shuffle function
shuffle(random_array);
test_array[]=random_array;

// assign members to groups
for(a=0;a<no_of_groups;a++)
          for(b=0;b<sizeof(test_array[a]);b++)
                              new_array[]=test[a][b];
for(c=0;c<no_of_groups;c++)
                    for(rank=c;rank<sizeof(new_array);rank+=no_of_groups)
                    group_array[c][rank]=new_array[rank];.

 add_group_data_into_Moodle:= mdl_groups_members < input(group_array
     [c][rank])

exit_session()
```

(d) Feed back platform (email and SMS)

To implement the feedback module, the clusters' data was linked to the feedback platform so that feedback can be disseminated based on the collaboration competence level. To provide feedback based on clustered data, an interface in Moodle which allows the instructor to SMS or email either the whole group in that cluster or select a single student in that cluster depending on whether the feedback is for the entire cluster or for a single student. For the SMS module, the Moodle interface was integrated with an API platform provided by Africa's Talking. For email, the server was configured to allow emails to be sent from the Moodle database.

(e) Prototype Testing

In order to test the prototype, a group task in the form of discussion forum in Moodle was given to a class of 36 students. The students were grouped randomly in groups of four and they were required to discuss online for a period of two weeks on the group task. After the two weeks, the forum statistics for each student in form of number of posts, number of replies and average ratings were generated, as shown in Table 3.9

Table 3.9: Forum statistics for 36 students

Student ID	Number of Posts	Number of Replies	Average Ratings
356	16	37	3
327	18	23	3
330	23	15	0
321	10	20	4
353	15	14	2
422	14	15	2
348	10	18	2
421	9	16	4
328	9	14	4
347	14	13	0
408	12	15	0
318	16	4	3
324	3	17	3
436	10	12	1
292	6	11	2
294	11	6	1
286	12	4	0
394	7	9	0
443	8	8	0
282	9	3	3
313	8	3	4
442	4	11	0
455	11	2	1
296	9	4	0
302	7	6	0
337	7	5	1
291	4	5	3
414	8	4	0
371	10	1	0
287	9	1	0
346	5	4	1
434	1	7	0
453	7	1	0
412	4	2	0
301	3	1	0
430	4	0	0

This forum data was used to test the following modules: (i) clustering, (ii) intelligent grouping, (iii) email and SMS

The testing of the above modules is discussed in details.

(i) Clustering

In order to establish whether the cluster execution in Moodle was working, firstly the forum statistics described in Table 3.9 were transformed into an Attribute-Relation File Format (ARFF) (*'testdata1.arff'*) file and executed in Weka software using SKmeans and EM clustering algorithms. Secondly, the same data was clustered using the same clustering algorithms using the clustering module in Moodle platform. Finally, through expert analysis, the two results from Weka and from Moodle were compared to ensure the clustering code in Moodle was perfect and gave the same result as in Weka.

In Weka software clustering using SKmeans and EM requires first a number of test to be run, and then establish the values of two important parameters (seed value and maximum alteration). In this study, cross validation was done with the test data for both SKmeans and EM in order to establish the best values for these parameters which could give results with high accuracy level. For Skmeans best seed value was 10 and maximum alteration value was 500, while for EM best seed value was 500 and maximum alteration value was 100. Using these parameters, the *'testdata1.arff'* file which contained the test data was run in Weka and the screen shot for the results of SKmeans and EM algorithms is shown in Figure 3.10 and 3.11 respectively.

```
=== Run information ===
Scheme:       weka.clusterers.SimpleKMeans -V -N 3 -A "weka.core.
EuclideanDistance -R first-last" -I 500 -S 10
Relation:     Book1_clustered-weka.filters.unsupervised.attribute.Remove-R1
Instances:    36
Attributes:   3
              post
              av_rating
              replies
Test mode:    evaluate on training data
=== Model and evaluation on training set ===
kMeans
======
Number of iterations: 3
Within cluster sum of squared errors: 2.598172668504364
Missing values globally replaced with mean/mode
Cluster centroids:
                                    Cluster#
Attribute         Full Data          0           1           2
                       (36)        (9)         (6)        (21)
========================================================
post
      Mean             9.25       8.2222          16      7.7619
      St. Dev.     +/-4.6866  +/-3.8006  +/-4.3359  +/-3.3898
av_rating
      Mean           9.1944    10.3333    20.3333      5.5238
       St. Dev.     +/-7.877  +/-6.7082  +/-8.8015  +/-4.3888
replies
        Mean         1.3056     3.3333           2      0.2381
       St. Dev.    +/-1.4894  +/-0.7071  +/-1.0954  +/-0.4364
Clustered Instances

0       9 ( 25%)
1       6 ( 17%)
2      21 ( 58%)
```

Figure 3.10: Testing results for SKmeans in Weka

```
=== Run information ===
Scheme:       weka.clusterers.EM -I 100 -N 3 -M 1.0E-6 -S 500
Relation:     Book1_clustered-weka.filters.unsupervised.attribute.Remove-R1
Instances:    36
Attributes:   3
              post
             av_rating
              replies
Test mode:    evaluate on training data
=== Model and evaluation on training set ===
EM
==
Number of clusters: 3
            Cluster
Attribute         0       1       2
               (10)    (18)    (22)
=====================================
post
  mean        14.5007   7.053  7.3142
  std. dev.     3.923  2.9868  2.5167
av_rating
  mean        16.0397  4.3844 11.1594
  std. dev.    8.4961  3.2656  6.2708
replies
  mean         1.6802  0.2143   3.323
  std. dev.    1.2499  0.4104  0.7291

Clustered Instances
0      10 ( 28%)
1      18 ( 50%)
2       8 ( 22%)

Log likelihood: -7.43506
```

Figure 3.11: Testing results for EM in Weka

For the SKmeans results in Figure 3.10, it can be observed that cluster 0 had 9 students (25%), cluster 1 had 6 students (17%) and cluster 2 had 21 students (21%). Cluster 0 is characterized by students who have higher mean value on the number of replies (3.33) than in the other two clusters. Cluster 1 is characterized by students who have higher mean value on the number of posts (16.00) and average rating for the posts (20.00) than the other two clusters. Cluster 2 is characterized by students who have low mean values on the number of posts (7.76), number of replies (5.52) and average rating for the posts (0.24) than the other two clusters.

For the EM results in Figure 3.11 it can be observed that cluster 0 had 10 students (28%), cluster 1 had 18 students (50%) and cluster 2 had 8 students (22%). Cluster 0 is characterized by students who have higher mean value on the number of posts (14.50) than in the other two clusters. Cluster 1 is characterized by students who have low mean value on the number of replies (0.21) and average rating for the posts (4.38) than the other two clusters. Cluster 2 is characterized by students who have high mean value on the number of replies (3.32) than the other two clusters.

Similar tests were done in Weka software using the two clustering algorithms (SKmeans and EM) and the same parameters for different forum statistics obtained from Moodle database. Table 3.10 summarizes the results of 3 different datasets where the number of students varied in the forum statistics.

Table 3.10: Summary results in Weka for SKmeans and EM in 3 different datasets

Total number of students	36		109		151	
Cluster Number	Skmeans	EM	Skmeans	EM	Skmeans	EM
0	9 (25%)	10 (28%)	12 (11%)	61 (56%)	35 (23%)	44 (29%)
1	6 (17%)	18 (50%)	39 (36%)	14 (13%)	35 (23%)	77 (51%)
2	21 (58%)	8 (22%)	58 (53%)	34 (31%)	81 (54%)	30 (20%)

From Table 3.10, we observe that both SKmeans and EM almost gave similar distribution pattens on the number of students in different clusters regardless of the total number of students involved. We find that in every set of data there is a cluster with high number of students and one with less number of students regardless of the type of clustering algorithm applied. However, this distribution patten does not correspond with cluster values for both algorithms. Through expert analysis, we found that cluster with low values had students who had a high number of posts, replies and average ratings. Therefore, ranking was required to be done before using the cluster results to determine the best students who can be assigned as group mentors in their groups.

In order to test the clustering module in Moodle, a custom *'Discussion'* block is created to view and manage the clustering algorithms from the Weka.PHP program. The custom block has the cluster option which can be accessed by the instructor in his course. The cluster option is supposed to load the Weka.PHP program which provides the user an interface for creating the clusters and a display form which loads the cluster instances with students' identities. In this study, the number of clusters created were three, which were supposed to be ranked to correspond to the three collaboration competence levels i.e. High, Medium and Low as described in Table 2.7. To create clusters, the instructor is required to load the Weka.PHP program and input the following:

1. The type of clustering algorithm by selecting from a drop down list

2. Number of clusters which should correspond to the number of collaboration competence levels. For this study the option was 3.

To confirm that the cluster module was working perfectly, the three data sets which were executed in Weka software were used to do clustering in Moodle and the two results were compared. The same parameters were also applied in both cases. This comparison involved firstly, comparing the number of students assigned in each cluster and determining whether they are the same in both cases. Secondly, visualizing the cluster assignment in Weka software for each student using the student identity and comparing it with the Moodle cluster output for each student and determining whether they correspond in both cases. For example, Figure 3.12 shows clustering results in Moodle for the first data set which had 36 students using SKmeans while Figure 3.13 shows clustering results in Moodle for the same data set using the EM algorithm. In both cases students' identities are used to output cluster assignment.

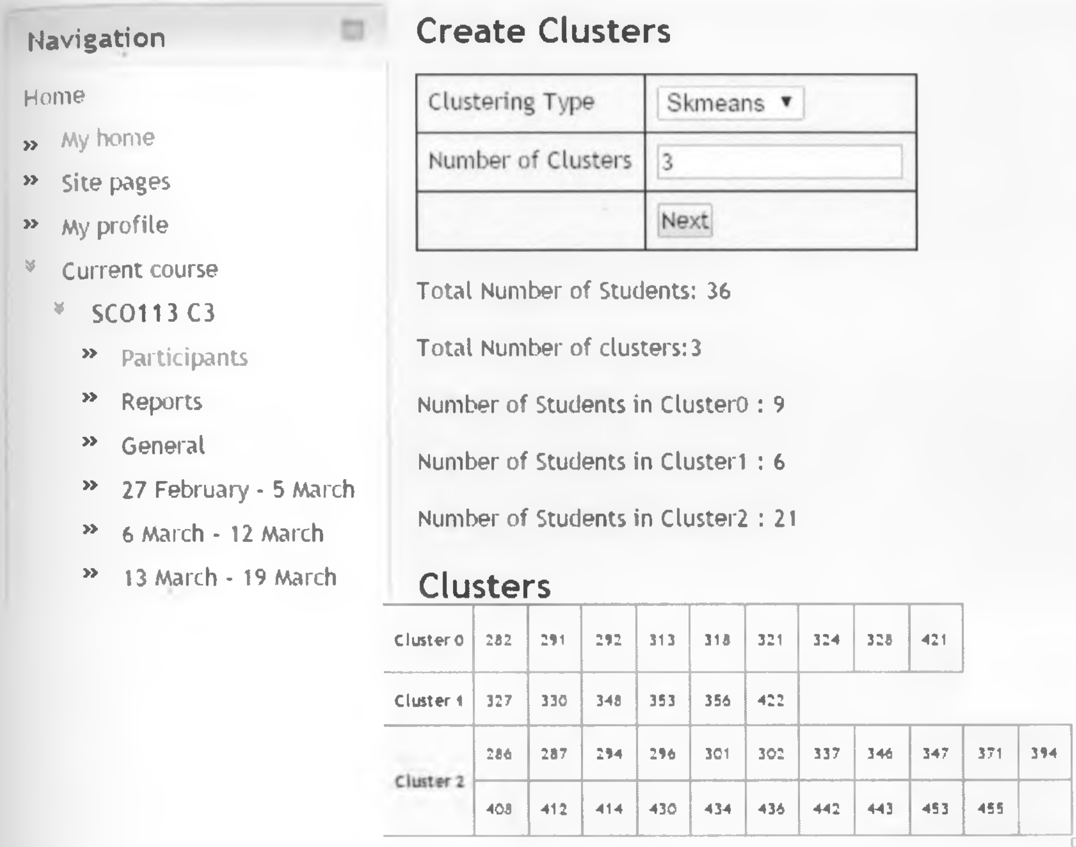

Figure 3.12: Testing results for SKmeans in Moodle

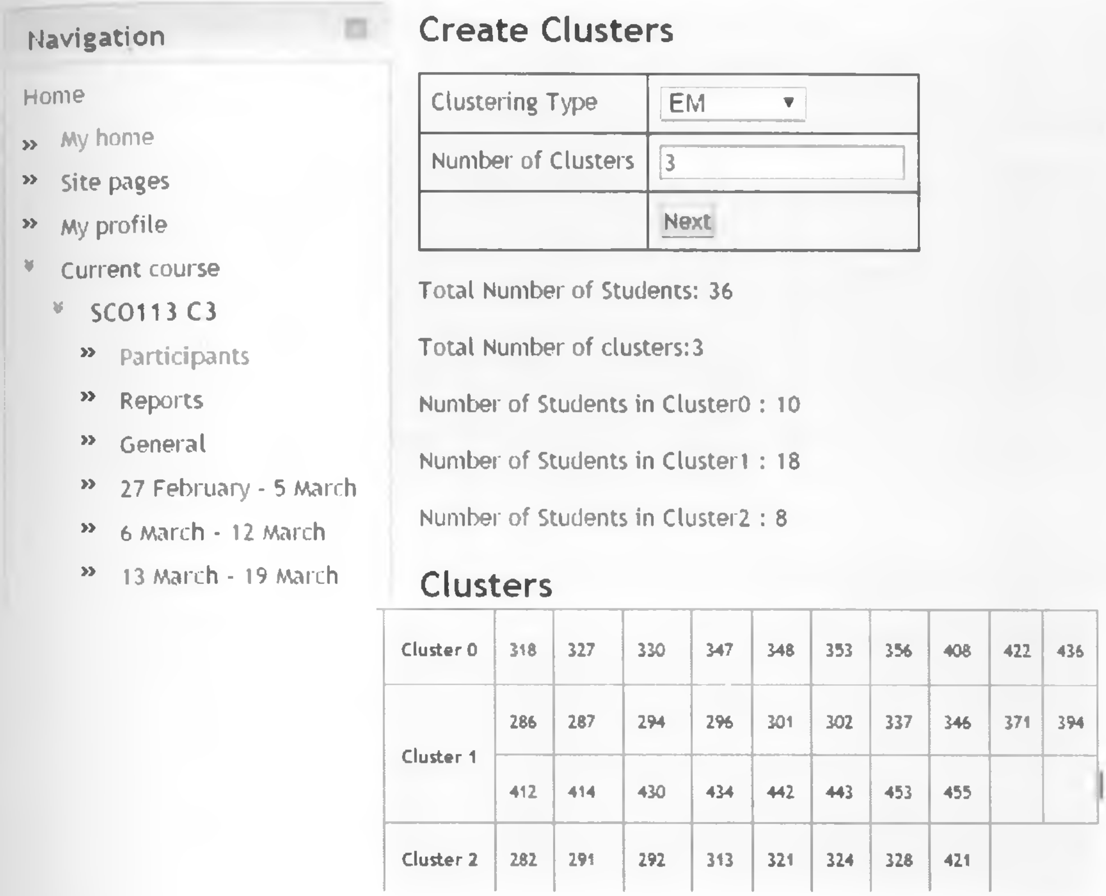

Figure 3.13: Testing results for EM in Moodle

The results from Moodle concurred with those obtained in Weka software both in terms of number students in each cluster and also in terms of cluster assignment for each student. From these results a comparative analysis was done to compare how students were being assigned to clusters by both algorithms in terms of the three attributes (number of posts, number of replies and average rating). Table 3.11 summarizes this comparative analysis for a data set of 36 students.

Table 3.11: Comparative analysis on cluster assignment

UserID	Number of posts (a)	Number of replies (b)	Average ratings (c)	Total points (a+b+c)	Cluster value (EM)	Cluster value (SKMeans)
356	16	37	3	56	cluster0	cluster1
327	18	23	3	44	cluster0	cluster1
330	23	15	0	38	cluster0	cluster1
321	10	20	4	34	cluster2	cluster0
353	15	14	2	31	cluster0	cluster1
422	14	15	2	31	cluster0	cluster1
348	10	18	2	30	cluster0	cluster1
421	9	16	4	29	cluster2	cluster0
328	9	14	4	27	cluster2	cluster0
347	14	13	0	27	cluster0	cluster2
408	12	15	0	27	cluster0	cluster2
318	16	4	3	23	cluster0	cluster0
324	3	17	3	23	cluster2	cluster0
436	10	12	1	23	cluster0	cluster2
292	6	11	2	19	cluster2	cluster0
294	11	6	1	18	cluster1	cluster2
286	12	4	0	16	cluster1	cluster2
394	7	9	0	16	cluster1	cluster2
443	8	8	0	16	cluster1	cluster2
282	9	3	3	15	cluster2	cluster0
313	8	3	4	15	cluster2	cluster0
442	4	11	0	15	cluster1	cluster2
455	11	2	1	14	cluster1	cluster2
296	9	4	0	13	cluster1	cluster2
302	7	6	0	13	cluster1	cluster2
337	7	5	1	13	cluster1	cluster2
291	4	5	3	12	cluster2	cluster0
414	8	4	0	12	cluster1	cluster2
371	10	1	0	11	cluster1	cluster2
287	9	1	0	10	cluster1	cluster2
346	5	4	1	10	cluster1	cluster2
434	1	7	0	8	cluster1	cluster2
453	7	1	0	8	cluster1	cluster2
412	4	2	0	6	cluster1	cluster2
301	3	1	0	4	cluster1	cluster2
430	4	0	0	4	cluster1	cluster2

From the table, it was observed that students who had high values in the three attributes were assigned in cluster 0 for EM while for SKmeans, majority were assigned in cluster 1. For those with low values majority were assigned in cluster 1 in EM and cluster 2 in SKMeans while majority of the average students were assigned cluster 2 in EM and cluster 0 in Skmeans. In summary, it was observed that majority of students who assigned cluster 0 in EM were being assigned cluster 1 in SKmeans, cluster 1 in EM were being assigned cluster 2 in SKMeans and those in cluster 2 in EM were being assigned cluster 0 in Skmeans. From this observation it was clear that the two algorithms had almost similar distribution of student to clusters. However, further empirical analysis is still necessary to sufficiently evaluate if this pattern can persist given different sizes of data sets.

(ii) Intelligent Grouping

To test the intelligent grouping module in Moodle, a custom block called *'IGLC Grouping'* menu is created to view, manage clustering and create groups using the intelligent algorithm. This custom block can be accessed by the Teacher role in his course. This custom block displays the option for creating clusters. After clustering the teacher can use the group option to create new custom groups by selecting the type of algorithm (SKMeans or EM), number of clusters, number of groups and then click on *'create groups'* button. After submission, the page shows the preview of cluster assignments and group assignments with user identities. The teacher can continue with this grouping by clicking on *'continue'*. It will replace the old groups and group members with new groups.

In order to assign group mentors, the clusters are first ranked and those students who are ranked highly in terms of forum statistics are regarded as highly collaborative and they will be assigned as mentors or group leaders in their respective groups. For example, in Table 3.11, it was found that in EM results majority of the highly collaborative students were assigned cluster 0. Therefore, when these results were used to create groups in Moodle, it was observed that the intelligent grouping algorithm was capable of picking students from different clusters and assign them to one group and at the same time assign students from cluster 0 as mentors in each group. This distribution creates heterogeneous groups based on learners' collaboration competence level. For example, the execution of the intelligent grouping algorithm was done on the results of EM clustering shown on Figure 3.13, where the number of groups were specified to be nine. The outcome was nine groups as illustrated in Figure 3.14.

Total Number of Students: 36

Total Number of clusters: 3

Number of Students in Cluster 0: 10

Number of Students in Cluster 1: 18

Number of Students in Cluster 2: 8

Clusters

Cluster 0	318	327	330	347	348	353	356	408	422	436
Cluster 1	286	287	294	296	301	302	337	346	371	394
	412	414	430	434	442	443	453	455		
Cluster 2	282	291	292	313	321	324	328	421		

Groups

Grouping(1)	347(Mentor)	318	371	337
Grouping(2)	330(Mentor)	286	414	328
Grouping(3)	422(Mentor)	296	301	282
Grouping(4)	353(Mentor)	443	346	313
Grouping(5)	348(Mentor)	430	287	321
Grouping(6)	327(Mentor)	412	453	324
Grouping(7)	436(Mentor)	434	302	292
Grouping(8)	356(Mentor)	455	294	291
Grouping(9)	408(Mentor)	394	442	421

Figure 3.14: Testing results from intelligent grouping algorithm

From this figure, it can be observed that the algorithm distributed the students in away such that each group is assigned four students who are members of different clusters hence, creating heterogeneous groups based on learners' collaboration competence level. In addition, Students who are in cluster 0 (highly collaborative cluster) are assigned a mentor role in their group membership as this cluster constitutes highly collaborative students as illustrated in Table 3.11. The term *'mentor'* indicated that they will play the mentor role during the discussion. Therefore, the testing confirmed that the intelligent grouping algorithm was capable of forming heterogeneous groups based on ranked clustered data.

After groups are formed, the instructor can go ahead and click the next button to move the group data into Moodle tables (*mdl_groups, mdl_groups_members*). These groups are availed in the grouping module in Moodle and the instructor can assign them to a discussion forum or any other group activity as desired. Figure 3.15 illustrates the final results which are presented to the instructor so that he can assign them to a discussion forum or any other group task. These results are loaded with student's profile details which include the actual student identities and names as per the student's enrollment data. The form also loads with an SMS and email interface on each group so that in case the instructor decides to communicate to the entire group or individual members, he can go ahead and communicate with his preferred mode. The next section describes how the communication platform works.

Discussion Group Members

Group1	munga_solomon_0749(Mentor), jacob_njogu_18403, michael_mwangi_3614, sang_maureen_4495	Mail	sms
Group2	annastacia_wangari_4493(Mentor), allan_chege_0764, abdiaziz_ibrahim_18092, kariuki_kelvin_4482	Mail	sms
Group3	andrew_mwangi_4471(Mentor), kenneth_munguti_0754, nandwa_moses_16868, kipkemoi_josphat_2409	Mail	sms
Group4	james_gideon_16896(Mentor), alexander_francis_14425, amatalo_mendeleeve_4510, marvin_nderitu_17599	Mail	sms
Group5	nyambura_jane_4458(Mentor), michelle_faith_4786, john_mbata_4500, joseph_otieno_1142	Mail	sms
Group6	imelda_auma_4503(Mentor), felix_ronoh_14426, mitchelle_ngaira_0758, thomas_amenya_4475	Mail	sms
Group7	julius_njuguna_0774(Mentor), william_macuga_15917, dennis_kiplangat_18163, john_kioi_4476	Mail	sms
Group8	ebrahim_mwendia_4658(Mentor), michael_muganda_0643, albert_ouma_4512, derrick_mbugua_13448	Mail	sms
Group9	caroline_chepkoech_0769(Mentor), mercy_mbithi_4492, josphat_wagura_17943, collins_omondi_0782	Mail	sms

Figure 3.15: Testing results on group assignment in Moodle based on intelligent grouping algorithm

(iii) Feedback platform

For the SMS and email services to work, students are required to provide their email and their mobile phone numbers when enrolling to a course or they can update their profiles in case they missed to provide the details during enrollment. Emails are recommended to be used when the instructor wants to provide lengthy feedback while SMS is recommended for short messages. For example, if an instructor wants to elaborate on how to improve a student's performance in the collaborative work, a mail service is recommended. However, if the instructor wants to pass a quick notification on the performance to a group or individual student, an SMS is recommended. For example, Figure 3.16. show a screen shot for sending SMS and emails based on clustered data where the cluster values have been ranked from most collaborative(cluster 0) to least collaborative (cluster 2) and student profile data (student identifier and names) has been availed in each cluster.

Clusters

Cluster0	mike_muuro_0333, 10737, patrick_onyango_10743, 11621, agnes_chepkemoi_7109, nthenya_kasanga_4282, judith_moraa_4267, 7091, 13127, james_kamau_13146 watiri_eva_0632, ndubai_gituuro_4286, wanjira_wamae_13879, 646, 17542, 4293, kennedy_wambua_13901, agnes_karonji_4287, njoroge_kimani_13840	Mail	sms
Cluster1	odhiambo_ouma_0652, muthui_musee_7093, kiprono_gideon_4272, njeri_gatutha_0640, ann_kandie_0657, josephine_chepkoech_4275, mwendwa_kyalo_4283, jonathan_lagat_4788, olonyi_abisa_10462, mwalimu_mulyungi_10686, dennis_muriithi_10703, chebet_patricia_10718 muteti_kyalo_10735, ochola_evans_10737, rotich_bett_11631, joe_kuria_13787, munene_kennedy_13884, salano_ombiri_14909, momanyi_mogaka_7091, kihia_wangari_7959, chelagat_bii_9072, gichuki_christine_11641,	Mail	sms
Cluster2	mathu_kimani_13847, erick_mugambi_9067 griffin_muteti_4281	Mail	sms

Figure 3.16: Testing results for sending cluster sms and emails

If the instructor wants to send a particular SMS or email to the whole cluster or group he is required to click directly on the email or SMS command for the specific group or cluster. Figure 3.17 shows a screen shot after clicking the email command and then the SMS command.

Figure 3.17: Testing results for sending an email or sms to a group of students in a particular group or cluster

For a particular student in a specific cluster or group, the instructor needs to click on the particular cluster or group to load an interface which allows the selection of an individual student. Figure 3.18 shows a screen shot after clicking on cluster 2. This form allows the instructor to select a single student from cluster 2 in order to send an individualized SMS or email. To confirm whether the email or SMS was delivered, a confirmation message is generated indicating the status of delivery.

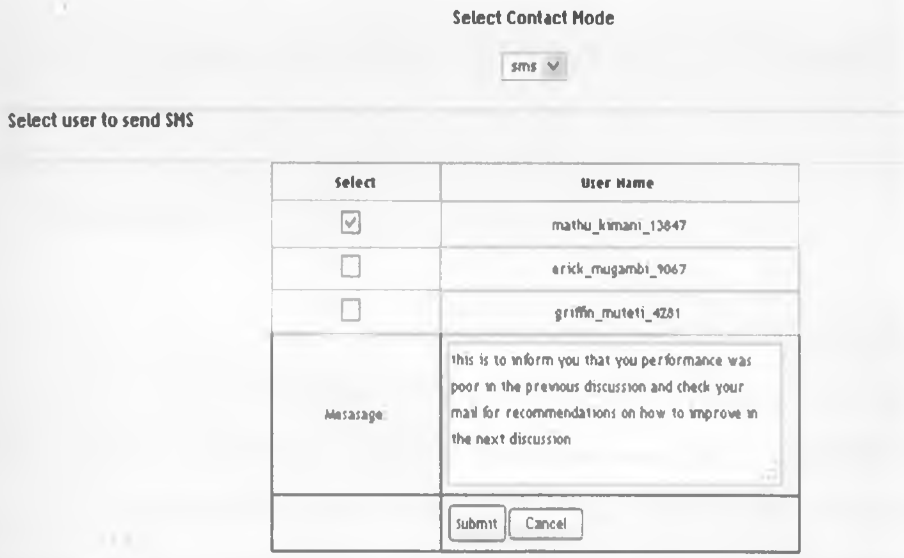

Figure 3.18: Testing results for sending individual SMS

3.4 Summary

This chapter has discussed how the system development methodology was carried out. Firstly, the chapter discusses how the pre-study was carried out to inform the system requirements process, system design and testing. Secondly, the the chapter has discussed how the system was designed and coded. The implementation of clustering algorithms to cluster students based on their collaboration competence level has also been discussed. The chapter also has discussed how intelligent grouping algorithm was implemented in order to form heterogeneous groups based on clustered data. The provision of customized feedback (SMS and emails) which is based on clustered data and groups has also been demonstrated. The testing process and test results for clustering algorithms, intelligent grouping algorithm and feedback platform has been discussed and demonstrated. The next chapter discusses how the system was evaluated through an experimental design methodology in order to validate it and evaluate its impact in group formation in an online collaborative learning environment.

CHAPTER FOUR

METHODOLOGY - EXPERIMENTAL DESIGN

4.1 Introduction

Group formation on group work has big impact on group performance. Depending on how the group is formed, it can result in homogeneity in student characteristics such that the peer learning is not effective. Thus, there is a need to constitute a heterogeneous group in collaborative learning which constitutes students with different collaborative competencies and knowledge levels. However, without empirical study it becomes difficult to conclude which group characteristics are desirable in the heterogeneity as different learning needs may require different group orientations. Previous research has focused on various group orientation techniques and their impact on group performance like different learning styles in group orientation (Alfonseca et al., 2006; Grigoriadou et al., 2006; Deibel, 2005). However, there is need to investigate the impact of other group orientation techniques on group performance like grouping students based on their Grade point Average (GPA), random assignment or based on their collaboration competence levels . Furthermore, most of the previous research in group formation lacks true experimental design methodology which is recommended when investigating learning outcomes with different instructional methods.

This research sought to investigate the impact of different group orientation techniques (GPA, Intelligent grouping , and Random) on group outcomes in an online collaborative learning environment. Hence, the research questions we intended to answer in this respect are:

1. Which group of learners amongst the intelligently grouped, randomly grouped and instructor grouped using GPA collaborates more effectively and performs better in an online group task?

2. What is the association between grouping method used and group outcomes in terms of: a) students' learning experiences; b) perceived problems; c) group leadership satisfaction and; d) group task satisfaction?;

3. What are the students' perceived benefits and challenges of online collaborative learning?

This chapter discusses the design of the experiment, the participants in the experiment, data collection procedure and data analysis methods.

4.2 Target Population

The students who participated in this study were first year students who were doing a Bachelor of Science in Computers Science and Bachelor of Science in Mathematics and Computer Science at Kenyatta University. First year students were targeted because senior students have socially interacted more and they do prefer to work through social groups which can skew the experiment results. These students were studying a first year course called *Foundations of Artificial Intelligence*. This is a course in computer science which has a number of topics like problem solving in a state space which has the potential to elicit some discussion, hence a good course to be done through collaborative learning. The entire population for the first year class was 108 students who had registered for the course by the time the research was being conducted. All the students were picked to participate in the study. Therefore, the sample size was same as the population.

The participants were enrolled into Moodle e-learning platform and each participant was provided with a user name and password. For the first two weeks, all the participants were trained on how to use Moodle e-learning platform. In the first week, a two hour tutorial was done twice to the participants to show them how to access the Moodle site, access the course notes, and engage in discussion forums and access quizzes or assignments.

After the training, the participants were randomly assigned into three classes with equal numbers (36 students per class). The randomization was done through generating random numbers in an excel worksheet. First, the registration numbers and names for the participants were entered into an excel worksheet and using the random function in excel a new column with random number was generated. Second, to make the random numbers values constants they were copied and pasted again as constant values. Third, the whole set of data was sorted with respect to the random values column. Fourth, students in the first 36 rows were assigned to class one, next 37 to 72 assigned to class two and lastly 73 to 108 were assigned to class three. A class size of 36 students was presumed appropriate to be under one instructor and also to assume normality in data analysis. Randomization was preferred as it has high chances of providing equal opportunity for participants to join different classes and ensure that the classes are equivalent. This also reduces the effect of extraneous variables such as subject characteristics which is major threat to internal validity (Fraenkel et al., 2012).

4.3 Research Design

A true experimental design was adopted where an experimental group and two control groups were used. The control groups played the role of comparison groups as they also received different treatment in terms of group orientation. Experimental design was adopted because it could help to identify the effect of independent variable (group orientation) to the dependent variable (group performance). The three classes which were formed through randomization as discussed earlier were used in the group design, where one class served as the experimental group and the other two classes as the control groups. The purpose of having two control groups was because it was observed from the pre-study two common methods for group orientation in collaborative learning dominated which were instructor based and random assignment. Thus, comparing these two methods with the intelligent grouping method was necessary. Each class was then assigned an instructor who was responsible to teach the course and oversee the discussions throughout the experiment period. The instruction design and teaching materials were prepared before the start of the course by the three instructors. This was to ensure same course materials and instruction design was used throughout in the three classes.

During the third and fourth week, students were given some discussion questions, such that for every week there was group task to be solved. Self-selected groups were used in all the three classes during this period of four weeks. The purpose of this discussion was to orient the students on forums in Moodle and at same time to generate discussion data which was to be used in the intelligent grouping. Self-selected grouping method was used because of: (i) known advantages such as allowing students to: communicate better, have positive attitude towards group work and feel more excited to work together Chapman et al. (2006) and (ii) to ensure internal validity as this grouping method was not included in the research question under study. At the end of four weeks discussion, the students did a pretest which was taken as the first Continuous Assessment Test (C.A.T). The pretest was also used to confirm whether the randomization method used in creating the three classes was heterogeneous in terms of learning capability.

During the sixth week, students were grouped into groups of four using different methods per each class. Group size of four was preferred as this was an average size which was small enough to represent heterogeneous learning characteristics and also to utilize the advantages that are realised when students discuss in groups of small size (Schellenberg, 1959). Students were expected to collaborate online at different times in the same place (same computer lab) using asynchronous communication tools . Each group had a group leader who was expected to initiate the discussion, moderate the discussion and summarize the main points. The following procedures were adopted to assign students into groups and also assign group leaders to each group:

1. In class one; the instructor used Grade Points Average (GPA) which was

calculated from the results for the last one semester. This class served as comparison group.

2. In class two; the instructor used the intelligent grouping algorithm to cluster students and group them based on learners' collaboration competence level. These collaboration competence levels were created using clustering algorithms and using the four weeks discussion data. This class served as an experimental group.

3. In class three; the instructor used random grouping method available in Moodle which automatically assigned students into groups of four. This class served as a comparison group.

After the exercise of grouping was over, students were informed about their groups, how the rest of the discussion was to be carried out and how the evaluation was to be done during the experiment period. Table 4.1 summarizes the research design procedures in terms of weekly schedule and corresponding events. Table 4.2 summarizes some of the threats to internal validity and how they were addressed in the research design.

Table 4.1: Summary on the experiment design procedure

Week	Event	Description	Remarks
3	Participants: 108 Students	1. Random assignment of 36 students to experiment group: Class one 2. Random assignment of 36 students to experiment group: Class two 3. Random assignment of 36 students to control group: Class three	Each class was assigned its own instructor and the same teaching materials and instruction design was used by all the three instructors
4,5	Discussion forums based on topic one and two	Through self selected groups students discussed questions which were drawn on the first topic of the course	Discussion forum was open throughout the two weeks
6	Pretest	A quiz with 30 Multiple choice questions given.	Multiple choice was preferred because of easy of marking and deep coverage of content
7	Group Orientations	1. Class one: GPA used to assign both group membership and group leader 2. Class two: Intelligent Grouping Algorithm both group membership and group leader 3. Class three: Random assignment both group membership and group leader	Same instructors continued with the classes
8 and 9	Discussion forums based on topic three	Students were given discussion questions two discussed as per their groups	Instructors monitored the discussions and advised students accordingly
10	Posttest and post study questionnaire	Students were given a posttest which had three tests. 1. Two hours discussion forum 2. A quiz which was multiple choice 3. Written test which was based on the 2 hours discussion After the posttest students were informed to respond to a questionnaire online.	Some few students did not participate for unknown reasons. The posttest was conducted in three labs such that each class had its own lab. This ensured that all the tests were done at the same-time by the three classes

Table 4.2: Summary of threats to internal validity

Type of Threat to Internal Validity	Measures taken
Subject characteristics	Randomization in assigning participants to groups and test (pretest) was done to measure the effectiveness of the randomization.
Location	Same learning environment was used, i.e. the whole experiment was conducted in KU
Instrumentation	Validation on each instrument was done as described in the respective sections and all tests were conducted the same time for all the groups. Different group was used to pretest the instruments rather than the participants. Successfully approved assessment tools in Moodle were used to assess the forums
Testing	Pretest and post test were different. Pretest was only meant to measure effectiveness of randomization.
Attitude of subjects	Student were informed about the purpose of the study at the start of the course and the tests were to part of the C.A.T for the course
Implementation	Three different instructors who are experts in the course were used to facilitate teaching of the course in the three classes but the same instructional materials were used throughout.

4.4 Instruments

The instruments which were used in this study include a pretest, posttests and a post study questionnaire. The section discusses how the instruments were constructed and the measures taken to ensure validity.

4.4.1 Pretest

Forty multiple choice questions were constructed where the question items were drawn from Artificial Intelligence (AI) books. The topic covered in the pretest was introduction to AI . To ensure the test involved thorough comprehension and critical thinking by the students, multiple choices were closely associated to the right answer for all items. The forty questions were then added into Moodle as a quiz and each question was assigned 1 mark.

Pretest validation

Before the pretest was given to the participants the following measures were taken to enhance validity:

1. All the three instructors were involved to provide different expertise when setting the questions and checking content validity.
2. Multiple choices were reshuffled dynamically by the system to avoid copying of answers among students.
3. A group of second year computer science students in KU, who were doing a similar course through Moodle, were given the pretest to do it online as written test

The following was observed:

- Some students had forgotten their passwords and user names for the Moodle e-learning platform even though the lecturer had provided the details at the start of the course as the students were required to be accessing the learning materials online and be participating in discussions forums. This was a clear indication some students had ignored the discussion forums and were only bothered on how to log in during the written test
- Slow internet connection delayed the exercise effecting the one hour time allocation for the pretest. To compensate for this, the affected students were given a second attempt to finish the remaining questions.
- 10 questions had a problem on multiple choice because when the multiple choices were reshuffled, a choice which included two choices was misleading as the choices could have been reshuffled and they still remained in the choice the way they were before the reshuffle. This necessitated the instructor to command the system to grade the 10 questions with zero marks reducing the final weight to 30 marks.

Scores from this pretest were within the normal curve. Therefore, the level of difficulty was acceptable under the normal distribution. The following was done before the participants were given the pretest:

1. Participant in the three classes were reminded to ensure the password and user names assigned to them, were operation and were able to access all instruction materials for the course before the pretest date.

2. Labs which had generators were preserved for the exercise in order to avoid cases of power blackout.

3. The pretest was reduced to thirty questions as the 10 questions which had the multiple choice questions were deleted from the initial list.

After these corrections were done, the three classes were given an individual test (pretest) which had 30 multiple choice questions. The test was posted in Moodle as quiz, each question had a weight of 1 mark and therefore the total score expected in the test was 30 marks. To minimize cross-over problem, all the participants did the pretest at the same time. Also, to minimize the problem of slow internet, the pretest was done between 8:00 am and 9:00 am as morning hours have less network traffic since most of the students in campus are busy reading in the library or attending classes. Having pretested the test and major problems minimized, the whole exercise was successful. The marking of the test was automated and students were informed about their scores immediately after logging off.

4.4.2 Posttest

The posttest was made up of three tests which were designed differently but the contents were drawn from the same topic. That way, different taxonomies on knowledge construction were examined as recommended in Bloom's taxonomy (Bloom, 1956). The first section was a discussion forum which required the students to solve state space search problem by discussing the following:

1. Description of the state space
2. Rules and operators for moving from one state to another
3. The possible solutions
4. Optimal solution and related heuristic function

State space search problems were preferred because:

1. They generate a lot of discussion since there could be multiple solutions depending on how the description of state space is given and the heuristic function used to generate solution.
2. It is possible to set many questions which are of the same weight by simply examining the four sections stated above.

For each class there were nine groups, where group size ranged between 3 and 4. To minimize crossover problem during discussion, nine questions of similar weight were constructed such that each group had its own question. But the nine questions were replicated in the three classes. However, the replication had no effect among the classes since each class was assigned a separate lab and the discussion forum was conducted the same time in all the three classes. Discussion forum was preferred because forum is a powerful tool in Moodle which allows course participants to post messages and reply to each other online.

The following assessment tools were used to mark the discussion forum:

1. **Rating tool in Moodle:** This is an assessment tool in Moodle which allows an instructor or a student to award a mark to a post (new post or a reply) in a discussion forum in form of rating. These ratings are then aggregated using the selected aggregate type to produce the final individual grade for that activity.

The instructor is required to set the maximum scale for the rates which is the maximum score and also the aggregate type in the forum settings. Different aggregation types do exist in Moodle which include: Average rating, Count of ratings, Maximum rating, Minimum rating and Sum of ratings. To demonstrate how these different rating methods work, we assume student X has posted three posts which have been rated out of 10 as follows:

post 1=10/10, post 2 = 3/10 and post 3 =5/10

2. **Average of ratings:** Provides the mean of all ratings to post in that forum. Then forum grade for student X will be (10+3+5)/ 3 =6/10 or 60%. This aggregation method is good when an instructor wants to evaluating the quality of each discussion post individually without worrying about quantity.

    (a) **Count of ratings:** Tallies the number of rated posts made by a user to define the final grade in the Forum. Each rating is equivalent to one point regardless of the value assigned in the rating. Then forum grade for student X will be (1+1+1) =3/10 or 30%. This aggregation method is good when the number of posts is important.

    (b) **Maximum rating:** Calculates the highest rating among all posts as the final grade. Hence once the highest grade is achieved in any post, it becomes the final grade. Then forum grade for student X will be 10/10 or 100%. This aggregation method is good for emphasizing the best work from participants, allowing them to post one, high-quality post as well as a number of more casual responses to others.

    (c) **Minimum rating:** Lowest rating of all posts is calculated as the student's final grade. Then forum grade for student X will be 3/10 or 30%. This aggregation method is good for emphasizing a culture of high quality for all posts. However, it also means that one poor response from a student will result in a low scoring activity grade.

(d) **Sum of ratings:** Adds each rating to calculate the activity grade, which cannot exceed the maximum scale for the Forum. Then forum grade for student X will be 10/10 or 100%. This aggregation method is good when an instructor wants to assess the quality and quantity of posts at the same time. However, the instructor should avoid rating an entry with the maximum value, because the student will automatically get 100% for the activity as it was the case for student X.

For this study, sum of ratings aggregate type was selected because of its capability to assess the quality and quantity of posts at the same time. A marking scheme prepared by the three instructors was adopted to achieve standard ratings. The maximum scale was set to 10 such that for a student to score the maximum, one was expected to post at least 10 different points. For a post to be awarded a score, the instructor had to check whether it is a solution to the problem or it demonstrates any synthesis of knowledge within the problem domain.

3. **Learning Analytics Enriched Rubric (LAe-R):** is an assessment rubric tool which contains "enriched" criteria and grading levels that are associated to data extracted from the analysis of learners' interaction and learning behavior in an online discussion forum. LAe-R has been developed as a plug-in for the Moodle learning management system and has been tested and proven to be very usable tool that is highly appreciated by teachers and students in evaluating online collaborative learning tasks (Dimopoulos et al., 2013). In forums, the tool analysis and visualizes data such as forum posts (new or reply messages), and number of files attached to the forum post. This tool was used to assess the quantity of posts sent by an individual in terms of log in, new post, replies and file attachment, therefore providing the assessment scores on how active a student was during the discussion period. The tool was preferred because it required minimal involvement of the instructor and included a number of parameters for assessing the individual participation level in the forum. This tool was downloaded

and installed in Moodle as plug-in and then integrated as an advanced assessment tool for the forum. Table 4.3 summarizes the parameters which were used to assign marks and the scaling levels. The scaling of marks on each parameter was discussed among the instructors and the final score was agreed as 10 marks. Table 4.3 describes the marking criteria adopted for the rubric analytic tool.

Table 4.3: Summary of parameters used to assign marks and assignment criteria

Parameter	Database object used in Moodle	Enrichment level check values	Marks/Points awarded	Maximal score
Number of occurrences/replies (p_1)	log and forum_posts	$p_1 >= 0$ $p_1 >= 1$ $p_1 >= 2$ $p_1 >= 3$	0 1 2 3	3
Number of files submitted into the forum (p_2)	forum_posts	$p_2 >= 1$ $p_2 >= 2$ $p_2 >= 3$	0 1 2	2
Number of new posts to the forum (p_3)	forum_posts	$p_3 >= 0$ $p_3 >= 1$ $p_3 >= 2$ $p_3 >= 3$	0 1 3 5	5

The rubric normalized score is calculated as:

$$G_s = \frac{\sum_{i=i}^{N} (g_i - min_i)}{\sum_{i=1}^{N} (max_i - min_i)} \tag{4.1}$$

Where g_i is the number of points given to the i-th criterion, min_i is the minimal possible number of points for the i-th criterion, max_i is the maximal possible number of points for the i-th criterion and N is the number of criteria in the rubric.

The second test was given inform of a quiz which consisted of 10 multiple choice questions which were constructed to examine the expected solutions in the discussion forum. This test was meant to measure individual's knowledge comprehension and knowledge construction during the discussion forum. The quiz was availed online immediately the discussion forum session was closed. Each student was given a single attempt for each item and was required to finish the 10 questions in the quiz within a period of 30 minutes. The process of marking and assigning scores for this quiz was automated, but students were not informed about their scores at this junction as they had to do another test. This was to avoid poorly scoring students being less motivated in the third test.

The third test was a written test which was constructed to test individual knowledge comprehension through short answers and easy questions. The test had weight of 20 marks and the tested items were based on the discussion forum. The test was administered immediately after the quiz and student were allocated one hour to do the test. Since the test was not meant to test memorization student were allowed to refer to their short notes they had prepared during the discussion session. This ensured that those students who had discussed a lot and arrived to the right solutions had a higher chance of scoring high if they prepared good notes from the discussion. The test was marked later using a marking scheme which was constructed by the three instructors and allocation of marks on each item was also agreed among the three instructors.

Posttest validation

Before the posttest was given to the participants the following measures were taken to enhance validity:

1. All the three instructors were involved to provide different expertise when setting the questions and checking content validity.
2. For the quiz, multiple choices were reshuffled dynamically by the system to avoid copying of answers among students.
3. A pretest on the post test was done with a group of second year computer science students in Kenyatta University who were doing a similar course through Moodle. It was found that most students were not able to handle the discussion questions,

majority requested for more examples in order to understand the concept. This prompted the instructors to organize for more classes to the participants in order to cover the same concepts with more examples and make it clear to avoid a similar situations as it was observed with the pretest group.

4.4.3 Posttest Questionnaire

The purpose of this questionnaire was to collect data on the students' experiences on the group task. These students' experiences were categorized into different categories as summarized in Table 4.4. Nineteen items in the questionnaire were close ended while three items were open ended. The Google doc. was used to construct the questionnaire, this made it easier to have the questionnaire availed online to the respondents.

Table 4.4: A summary of the questionnaire items

Item Number	Type	Information Gathered
Items 1-6	Multiple choice	Demographic information which included the email address, gender, group, class, frequently used tool to communicate online, previous knowledge on Moodle.
Items 7-10	Multiple choice	Problems experienced when doing group task
Item 11	5 Point likert scale	Whether the group task helped the individual learner to learn the tested concepts
Item 12	Yes/No	Who was a group leader and non leader
Items 13 &15	5 Point likert scale	Self evaluation on how effective the group leader was in leading the group
Item 16	5 Point likert scale	Whether the group leader played an effective role in leading the group
Item 17	Yes/No	Those who were not comfortable to continue with their group membership and those who were comfortable
Item 18	Short answer	Reasons for the choice provided in number 17
Item 19	5 Point likert scale	Collaboration experiences among the members in their group membership
Item 20	Open ended	Students' best experiences during the group activity
Item 21	Open ended	Students' worst experiences during the group activity
Item 22	Open ended	Students suggestions on how to improve the online discussion

Validation of the instrument

To ensure validity, content related evidence was used and two experts in e-learning were requested to review the content and the format of the questionnaire. The following comments were given by the two experts:

1. Add administrative details such as identification, school name, course unit, whether it is a CAT or not, how long it should take, and that they are free to participate (ethics)

2. Write down all the relevant aspects of interaction which you are interested in this study (some sort of framework) before you develop questions. Then write questions to cover all of them. Or if you have the questions, use the list to check the questions to ensure all the important aspects of interaction are covered. Then ask if they are satisfied with those specific aspects of interaction – break it down to get more valid information. Do the same for online collaborative learning experiences

3. Add an item to rate leadership skills

4. Add an item on student best experiences

Based on their comments some of the items were rephrased, more items were added, some content enriched and reformatting done as recommended. Content-related evidence was adopted to ensure the instrument contained adequate sample of the students' experiences on online collaborative learning as observed from the pre-study. The questionnaire was also pretested with a group of second year computer science students in KU, who were doing a similar course through Moodle. About fifty students were selected and emailed the questionnaires that were completed online. The Cronbach's coefficient alpha for the 5-point likert scale items had satisfactory reliability (alpha=0.86) (Nunnally, 1978).

4.5 Data Collection and Analysis

The posttest was done in three phases. During the first phase, the students were given a discussion question which was posted as forum in Moodle to discuss online for a period of three hours for all the three classes at the same time but in different labs. Separate groups were used in Moodle to ensure student only discussed within the group membership they were assigned. The discussion was later marked using two assessment tools which included instructor rating of the posts and rubric analytical tool which was customized to automatically assign marks according to the level of student's participation in terms of number of posts and number of replies to a post.

During the second phase, the students were given an individual quiz which had 10 multiple choice questions related to the discussion forum they had. The quiz was posted online in the Moodle and was attempted by all the three classes the same time within a period of 30 minute. It was then automatically marked out of 10, which was the maximum score.

During the last phase, the students were examined with a written test which had short answer questions. The total marks for this exam was 20. Answers to these questions were supposed to be obtained from the summary of the discussion. This was aimed at testing the student's level of knowledge construct in the domain area under discussion. The pretest and post test results were archived in Moodle database.

The posttest results were analyzed through SPSS in order to answer the following hypothesis:

- Null Hypothesis: There is no significant difference in mean scores between the three classes
- Alternative hypothesis: There is significant difference in mean score between the three classes

Both descriptive and One-Way ANOVA among the groups were done to answer the above hypothesis.

For the posttest questionnaire, a total of 108 students in the three classes were emailed the final questionnaire. Participants accessed the questionnaires items online and participation was voluntary. A total of 90 students responded - 83% response rate which was considered adequate for analysis. The collected data was exported from Moodle database to SPSS and coded in order to carry out both descriptive and inferential statistics as per the research objectives. Using SPSS, quantitative analysis was carried out, and the results were tabulated. For the open-ended items coding was done based on different themes observed from the results. To compare the students' experiences with different group formation methods, cross-tabulations were carried out on various items as per the research questions.

4.6 Summary

This chapter has discussed how the true experimental design was carried out. Due to the nature of this study, two comparison groups were used as control groups. Randomization was used to create three groups where one was an experimental group and the other two were comparison groups which served as control groups. A pretest was given after randomization to check if the groups were equivalent in terms of knowledge level. In the experimental group, the students were assigned to group membership through the intelligent grouping algorithm and this group was labeled as class two. In the comparison groups, participants in one of the group were assigned to group membership through random selection simply by using the default grouping method in Moodle. This class was labeled as class one. The other comparison group, GPA for participants was calculated and student were ranked and randomization was done to form heterogeneous groups such that in each group there was one student with higher, average and low GPA value. This class was labeled as class three. Three instruments were constructed to measure the group performance. These included the quiz, discussion forum and written test. Thereafter, a posttest questionnaire was constructed to collect data on student experiences during the group activity in an online collaborative learning environment. All the instruments were validated through expert analysis and pretesting with students who were doing a similar course but at a different level.

CHAPTER FIVE

RESULTS AND DISCUSSION

5.1 Experimental Results

5.1.1 Introduction

The purpose of this study was to investigate the effectiveness of different group orientations on group collaboration learning tasks and outcomes in an online collaborative learning environment. This experiment was conducted in order to answer the following three research questions:

- RQ_1: Which group of learners amongst the intelligently grouped, randomly grouped and instructor grouped using GPA performs better in an online collaborative learning environment?

- RQ_2: What is the association between grouping method used and group outcomes in terms of: a) students' learning experiences; b) perceived problems; c) group leadership satisfaction and; d) group task satisfaction?

- RQ_3: What are the students' perceived benefits and challenges of online collaborative learning?

This chapter presents the results obtained from the experimental study and discusses them based on the above research questions.

5.1.2 Pretest Results

Table 5.1 summarizes the descriptive statistics for the pretest which was done by the three classes. A total of 108 students participated in the pretest, where 36 students were in class one, 35 student in class two and 33 students in class three. The mean scores for class one, two and three was 15.5278, 17.3939, 16.1286 while the standard deviation was 3.4230, 2.7720, 3.6428 respectively. This shows that there was slight difference between the means in the three classes. Therefore it was necessary to carry out One-Way ANOVA test to determine whether there was any statistically significant difference among means. Table 5.2 summarizes the results of One-Way ANOVA test. The p-value obtained was 0.064 which is above the required .05 alpha level. Therefore, there was no statistically significant difference between the three means and multiple comparisons through post-hoc analysis was not necessary. With no statistical difference observed from the means, it was evident that, the three classes were homogenous and therefore randomization assignment of participants to different classes was effective.

Table 5.1: Descriptive statistics for the pretest

Class	N	Mean	95% Confidence Interval for Mean		Min	Max
			Lower Bound	Upper Bound		
Class One	36	15.5278	14.3696	16.686	10	24.5
Class Two	33	17.3939	16.411	18.3768	11.5	23.5
Class Three	35	16.1286	14.8772	17.3799	8.5	25.75

Table 5.2: One-Way ANOVA analysis for the pretest

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	61.937	2	30.968	2.825	0.064
Within Groups	1107.147	101	10.962		

Df is Degree of freedom; F is value to determine whether the results are significantly different; Min is the minimum score, Max is the maximum score, N is the number of participants and Sig. is the value to be compared with the alpha value (0.05).

5.1.3 Posttest Results

Table 5.3 summarizes the descriptive statistics for the forum rating scores which was done by the instructor in the three classes. A total of 100 students participated in the discussion forum, where 33 students were in class one, 33 student in class two and 34 students in class three. The maximum score expected from the instructor's rating was 10 points. The scores ranged between 1 to 10 points. Class two produced the minimum score of 1 point which could have dragged down the mean score as the median score for class two was higher than that of class three. As observed from the modal values, majority scored 10 points in all the three classes. The mean scores for class one, two and three were 7.6364, 7.2727 and 7.7647. This shows that there was slight difference between the mean scores in the three classes. However, the ANOVA analysis results shown in Table 5.4 indicate significance value is 0.731($p = .731$), which is above the alpha value (0.05). Therefore, there was no statistically significant difference in the mean score for the forum ratings between the three classes.

Table 5.3: Descriptive statistics for the forum ratings

	N	Mean	95% Confidence Interval for Mean		Min	Max	Median	Mode
			Lower Bound	Upper Bound				
Class One	33	7.6364	6.6906	8.5821	3	10	8.5	10
Class Two	33	7.2727	6.2455	8.2999	1	10	7	10
Class Three	34	7.7647	6.9632	8.5662	3	10	6	10

Table 5.4: One-Way ANOVA analysis for the Forum ratings

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.341	2	2.17	0.314	0.731
Within Groups	670.299	97	6.91		

Table 5.5 summarizes the descriptive statistics for the forum Rubric Analytical assessment tool. A total of 100 students participated in the discussion forum, where 33 students were in class one, 33 students in class two and 34 students in class three. The maximum score expected from the rubric analytical tool was 10 marks. The scores ranged between 0 to 9 points. This was an indication that some students neither sent a post or did they reply, hence scoring 0 points. This happened in all the three classes. Class two had the lowest value for mode which was 7 marks but the median value was

higher than class three but lower than that of class one. The mean scores for class one, two and three were 5.5758, 5.0909 and 5.4412 respectively. This shows that there was slight difference between the mean scores in the three classes. However, ANOVA analysis results shown in Table 5.6 indicate significance value is 0.771 ($p = .771$), which is above the alpha value (0.05). Therefore, there was no statistically significant difference in the mean score for the rubric analytical scores between the three classes.

Table 5.5: Descriptive statistics for the Rubric Analytic tool

	N	Mean	95% Confidence Interval for Mean		Min	Max	Median	Mode
			Lower Bound	Upper Bound				
Class One	33	5.5758	4.5806	6.5709	0	8	6.5	8
Class Two	33	5.0909	4.1143	6.0675	0	9	6	7
Class Three	34	5.4412	4.4351	6.4473	0	9	5.5	8

Table 5.6: One-Way ANOVA analysis for the Rubric Analytic tool

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.14	2	2.07	0.261	0.771
Within Groups	769.17	97	7.93		

Table 5.7 summarizes the descriptive statistics for the forum scores (Instructor's rating and rubric analytical tool scores) which was done by the three classes. A total of 100 students participated in the discussion forum, where 33 students were in class one, 33 students in class two and 34 students in class three. The maximum score expected from the instructor's rating was 10 points and the same for rubric analytical tool. Hence the total expected maximum score was 20 points. The scores ranged between 1 to 19 points. The mean scores for class one, two and three were 13.2121, 12.3636 and 13.2059 respectively. This shows that there was a slight difference between the mean scores in the three classes. However, ANOVA analysis results shown in Table 5.8 indicate that the significance value is 0.727 ($p = .727$), which is above the alpha value (0.05). Therefore, there was no statistically significant difference in the mean score for the forum ratings between the three classes.

Table 5.7: Descriptive statistics for the Forum scores

	N	Mean	95% Confidence Interval for Mean		Min	Max
			Lower Bound	Upper Bound		
Class One	33	13.2121	11.4598	14.9644	4	18
Class Two	33	12.3636	10.5604	14.1669	1	19
Class Three	34	13.2059	11.5079	14.9039	3	19

Table 5.8: One-Way ANOVA analysis for the Forum ratings

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15.8	2	7.9	0.321	0.727
Within Groups	2390.71	97	24.646		

Table 5.9 summarizes the descriptive statistics for the quiz scores which was done by the three classes. A total of 97 students did the quiz, where 34 students were in class one, 31 students in class two and 32 students in class three. The quiz was marked out of 10 and the scores ranged between 2.75 to 10 points. Majority performed well as observed from the modal value which was 10 marks. The mean scores for class one, two and three were 7.5074, 8.2742 and 8.0547 respectively. Class two had the highest mean score (8.2742) and median score (9.00). This shows that there was a slight difference between the mean scores in the three classes. However, ANOVA analysis results shown in Table 5.10 indicate significance value is 0.247 ($p = .247$), which is above the alpha value (0.05). Therefore, there was no statistically significant difference in the mean score for the quiz between the three classes.

Table 5.9: Descriptive statistics for the Quiz

	N	Mean	95% Confidence Interval for Mean		Min	Max	Median	Mode
			Lower Bound	Upper Bound				
Class One	34	7.5074	6.7449	8.2698	4.25	10.0	7.25	10.00
Class Two	31	8.2742	7.6454	8.9030	2.75	10.0	9.00	9.50
Class Three	32	8.0547	7.4282	8.6811	4.75	10.0	8.75	10.00

Table 5.10: One-Way ANOVA analysis for the Quiz

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.243	2	5.121	1.419	0.247
Within Groups	339.322	94	3.61		

Table 5.11 summarizes the descriptive statistics for the scores in the written test which was done by the three classes. A total of 97 students did the written test, where 33 students were in class one, 31 students in class two and 33 students in class three. The written test was marked out of 20 and the scores ranged between 2.0 to 17.0 points. The mean scores for class one, two and three were 7.0606, 7.7419 and 8.3636 respectively. Both the mean and median score for class two was higher than those of class one but lower than class three. This shows that there was a slight difference between the mean scores in the three classes. However, ANOVA analysis results shown in Table 5.12 indicate significance value is 0.321 ($p = .321$), which is above the alpha value (0.05). Therefore, there was no statistically significant difference in the mean sore for the written test between the three classes.

Table 5.11: Descriptive statistics for the written test

	N	Mean	95% Confidence Interval for Mean		Min	Max	Median	Mode
			Lower Bound	Upper Bound				
Class One	33	7.0606	5.8166	8.3046	2	16	6	6
Class Two	31	7.7419	6.5744	8.9095	3	17	7	6
Class Three	33	8.3636	7.0374	9.6898	2	16	8	10

Table 5.12: One-Way ANOVA analysis for the written test

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	28.034	2	14.017	1.15	0.321
Within Groups	1145.451	94	12.186		

Table 5.13 summarizes the descriptive statistics for the total scores in the three post test

which were done by the three classes. Results of 100 students are shown, where 34 students were in class one, 33 students in class two and 34 students in class three. The total score was out of 50 and the scores ranged between 3.75 to 43.50. The mean scores for class one, two and three were 26.7941, 27.4091 and 28.9044 respectively. Class two had the majority scoring the highest (29 points) even though the median and the mean score went slightly lower compared with the other two classes. This shows that there was a slight difference between the means in the three classes. However, ANOVA analysis results shown in Table 5.14 indicate significance value is 0.564 ($p = .564$), which is above the alpha value (0.05). Therefore, there was no statistically significant difference in the mean score on the total score for posttest between the three classes.

Table 5.13: Descriptive statistics for the total Scores (Posttest)

	N	Mean	95% Confidence Interval for Mean		Min	Max	Median	Mode
			Lower Bound	Upper Bound				
Class One	34	26.7941	23.69	29.8982	9.3	43.5	28.12	26.5
Class Two	33	27.4091	24.7593	30.0588	3.8	42.25	27.37	29
Class Three	34	28.9044	25.9255	31.8833	13	43.25	27.62	23

Table 5.14: One-Way ANOVA analysis for the total Scores (Posttest)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	80.012	2	40.006	0.576	0.564
Within Groups	6804.163	98	69.43		

5.1.4 Results on Posttest Questionnaire

(a) Participants' Demographic Information

A total of 90 students responded out of 108 students who had participated in the study, with class one having 29, class two 29 and class three 32. There was a big gap for the gender equity as 75% were male and 17% female. The low percentage for female participants was expected because the study was based on students who were doing computer science course which had few female students enrolled for the course. Short Message Service (SMS) was reported as the most preferred communication tool (69%), followed by social media (Facebook and Twitter) (22%) and the least commonly used were email (7%) and phone calls (2%) in an online collaborative learning environment. 52% had previous knowledge on how to use Moodle e-learning platform while for 48% of the students, it was their first experience in using Moodle. Table 5.15 summarizes the demographic information.

Table 5.15: Summary on demographic information

Demographic Information (n=90)		Frequency	Percentage
1. Gender	Male	75	83%
	Female	15	17%
2. Class	Class one	29	32%
	Class two	29	32%
	Class three	32	36%
3. Frequent of use on communication tools	email	6	7%
	SMS	62	69%
	Socila Media	20	22%
	Phone Calls	2	2%
4. Previous knowledge on how to use Moodle	Yes	47	52%
	No	43	48%

(b) Problems Experienced During the Group Task

Students were provided with a list of six problems which had been identified from the pre-study and they were requested to identify from the list the ones they experienced. From the results shown in Table 5.16 two major problems were observed. Firstly, 47 participants (52%) reported that there was individual contribution imbalance with some members contributing less than others. Secondly, 43 participants (48%) reported that there was lack of feedback on participation among peers. The rest of the problems included: lack of coordination from group leader (22%), problems with negotiation skills such that it was difficult to agree on a common goal (17%), conflict and problems in reaching consensus in the group exercise (12%) and posting of irrelevant comments by members (3%). Table 5.16 and Figure 5.1 summarizes the frequencies on the observed problems in terms of mean in the three groups together with the overall mean. Participants who experienced the problems of lack of participation feedback, individual contribution imbalance and problems with negotiation skills were fewer in class two than the other two classes. However, as observed from p-values, there was no statistical significance on the difference among the three classes.

Table 5.16: Summary of frequencies on the observed problems

Problems experienced during the group task	Mean				
	Overall (n=70)	Class one (n=29)	Class two (n=29)	Class three (n=32)	*p – value*
Lack of participation feedback			0.48	0.44	0.822
Conflict and problems in reaching consensus in the group exercise	0.12	0.14	0.14	0.09	0.829
Individual contribution imbalance with some members contributing less than others	0.52	0.66	0.41	0.5	0.175
Problems with negotiation skills such that it was difficult to agree on a common goal	0.17	0.24	0.1	0.16	0.363
Lack of coordination from Group Leader	0.22	0.1	0.28	0.28	0.174
Posting of irrelevant comments by members	0.03	0	0.03	0.06	0.397

The mean is equivalent to the proportion of yes responses in the above table
*P-value: Significance of difference between class one, class two and class three: *P <* 0.05

Figure 5.1: Problems experienced during the group task

(c) Effectiveness of the Group Discussion in Learning the Concept

Students were asked to indicate how effective the group discussion was in learning the tested concepts in a 5 point likert scale (Strongly agree, Agree, Neutral, Disagree and strongly disagree) on six key elements which included; (i) how easy it was to learn through the discussion forums in Moodle; (ii) how participation in group activity improved his understanding of the topic under discussion; (iii) whether the group learning was more effective than individual learning; (iv) whether contribution from others helped the individual student to have better understanding of the problem solving concept in AI; (v) whether new posts to the forum provided the learner with a new perspective of the topic in discussion and; (vi) whether the online discussion forum was better than face to face learning of the concept. As summarized in Table 5.17, all the elements were positively rated with some having very high mean responses above 4.0. Mean values for class three were higher than for the other two classes for all items.

Table 5.17: Mean values for responses on the effectiveness of the group discussion as a learning tool

Outcome measure on the effectiveness of the group discussion as a learning tool	Mean				*p – value*
	Overall (n=70)	Class one (n=29)	Class two (n=29)	Class three (n=32)	
I found it easy to learn through the Discussion Forums in Moodle	4.04	3.9	4.07	4.16	0.629
By reading the contribution of others I had a better understanding of the problem solving concept in AI	4.04	4.07	3.83	4.22	0.155
The participation in Group Activity improved my understanding on the topic under discussion	4.02	3.93	4.03	4.09	0.573
When group members created new post it provided me with a new perspective of the topic in discussion	3.98	3.79	3.9	4.22	0.49
I learnt more about the subject matter under discussion in the group exercise than I would if I worked individually	3.88	3.72	3.86	4.03	0.57
In online discussion forums I learnt more than discussions in other face to face (Lecture) methods	3.42	3.17	3.48	3.59	0.21

Rating are based on a 5-point likert scale where 1 = strongly disagree and 5 = strongly agree

*P-value: Significance of difference between class one, class two and class three: *P <* 0.05

(d) Group Leader Experiences on Group Leadership

Out of 90 participants, 23 played the role of group leaders. Table 5.18 summarizes the respondents mean values on 5 point likert scale on three items. Majority of them were satisfied with their roles and were motivated to read widely as observed from the mean values which were above 4.00. As shown in Figure 5.2, 14 group leaders (6 from class one, 5 from class two and 3 from class three) found it easy to read their groups while 6 (2 from class two and 4 from class three) of them found it hard to lead their groups.

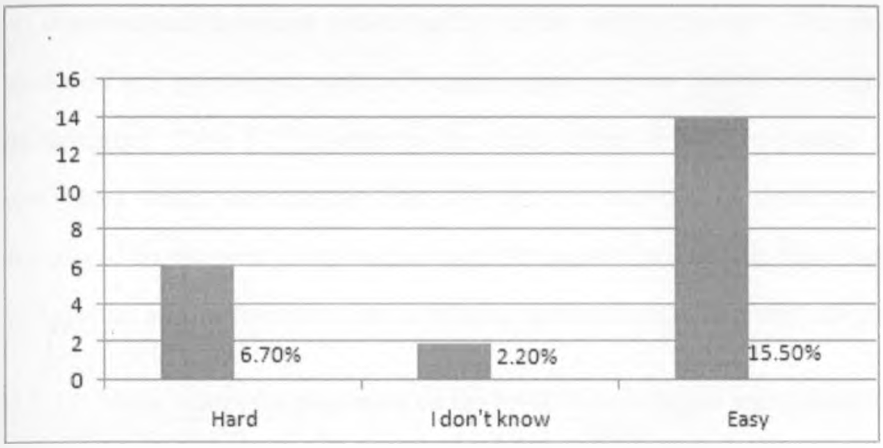

Figure 5.2: Rating of the group leadership task

Table 5.18: Mean values for responses on group leaders' perceptions in leading the group

Outcome measure on group leaders' experiences on group leadership	Mean				*p – value*
	Overall (n=70)	Class one (n=29)	Class two (n=29)	Class three (n=32)	
The responsibility motivated me to read widely enabling me to lead other members in the group	4.27	4.14	4.38	4.29	0.427
I enjoyed playing the role of a Leader in my group	4	4.14	4	3.86	0.461
Members in my group were reluctant to contribute	2.36	2.57	2.5	2	0.6

Rating are based on a 5-point likert scale where 1 = strongly disagree and 5 = strongly agree
*P-value: Significance of difference between class one, class two and class three: *P <* 0.05

(e) Group Members Satisfaction with their Group Leader

Students were asked to indicate how effective the group leader was in coordinating group members during the group task by responding to five key elements on group leadership, which included: (i) enjoyed working with their group leader; (ii) coordinated the group exercise well and kept the group on-track (kept the group focused and organized); (iii) summarized the group's discussion and came up with the conclusions; (iv) managed conflict and differences of opinions within the group task and; (v) demonstrated thorough understanding of the subject content. The responses were given on a 5 point-likert scale (Strongly agree, Agree, Neutral, Disagree and strongly disagree). Table 5.19 summaries the mean values of these responses. As per the mean values which were ranging from 2.88 to 3.52, students positively recognized the roles played by the peer group leaders with the highest ranked role being enjoying working together and the lowest ranked role being summarization of group's discussion.

Table 5.19: Mean values for responses on the level of satisfaction with group leader

Outcome measure on the effectiveness of group leader	Mean				
	Overall (n=56)	Class one (n=17)	Class two (n=19)	Class three (n=20)	*p – value*
I enjoyed working with my group leader	3.52	3.59	3.42	3.55	0.422
Our group leader coordinated the group exercise well and kept the group on-track- kept the group focused and organized	3.21	3.18	3.11	3.35	0.299
Our group leader demonstrated thorough understanding of the subject content	3.2	3.18	3.16	3.25	0.295
Our group leader managed conflict and differences of opinions within the group task	3.09	2.88	3.16	3.2	0.248
Our group leader summarized the group's discussion and came up with the conclusions	2.88	2.47	3	3.1	0.020*

Rating are based on a 5-point likert scale where 1 = strongly disagree and 5 = strongly agree
P-value: Significance of difference between class one, class two and class three: $*P < 0.05$

Out of 78 respondents, 64% reported they were willing to continue with the same members in their group if given another group task, while 36% recommended for a change. Reasons for those who were willing to continue and those who were not are summarized in Table 5.20 and Table 5.21 respectively. Those who opted they would like to continue, majority of them said that their members contributed well (29%) and their discussion was lively (12%). Those who opted they would not like to continue, majority of them said that they would like to get new experiences from new members

(10%) while others sited lack of participation from members in their group (6%).

Table 5.20: Summary on reasons for willing to continue with the same group membership

Reasons for willing to continue	Frequency	percentage	Class one	Class two	Class three
Members Contributed well and were able to come up with a solution	29	32.20%	9	9	11
Group members were active, and discussion was lively and enjoyable	12	13.30%	3	7	2
Our leader coordinated the discussion topics well	5	5.60%	0	3	2

Table 5.21: Summary on reasons for students not willing to continue with the same group membership

Reasons for not willing to remain	Frequency	percentage	Class one	Class two	Class three
To get new experience and exposure with new members	10	11.10%	3	4	3
Lack of participation among other member of the group	6	6.70%	4	2	0
Interaction was not good	4	4.40%	1	1	2
Lack of coordination from group leader	4	4.40%	1	1	2
To have more active members	2	2.20%	1	0	1

(f) Group Task Satisfaction Level

Students were asked to indicate the level of satisfaction on different elements of the group activity. The responses were given in a 5-point likert scale (Strongly agree, Agree, Neutral, Disagree and Strongly disagree). The mean values for the responses are summarized in Table 5.22. Students positively rated the items with some having a mean value above 4.00 and they highly agreed they would like to have more discussion forums in future.

Table 5.22: Group task satisfaction level

Outcome measure on group task satisfaction	Mean				
	Overall (n=78)	Class one (n=24)	Class two (n=27)	Class three (n=27)	*p – value*
I would recommend online discussion forums in future studies in my course work	4.28	4.12	4.47	4.33	0.461
I think all our group members were given fair opportunity to contribute	4.19	3.96	4.33	4.26	0.455
I would recommend for more group activities with my group members	4	3.72	4.11	4.15	0.467
I enjoyed working with my peers in our group activity	3.99	3.71	4.11	4.11	0.467
The group size was optimum for effective discussion	3.96	3.79	4.22	3.85	0.548
In my group activity, members were free to critiqueize each other contribution in a positive and constructive manner	3.95	3.92	4.04	3.89	0.574
Time allocated was enough to complete the group activity	3.83	3.46	4	4	0.348
In our group activity, I was able to negotiate with my peers and reach to a consensus	3.76	3.75	3.78	3.74	0.739
I was satisfied with the level of contact I had with my peers	3.6	3.5	3.56	3.74	0.086
One or two members dominated the group exercise	2.76	2.96	2.81	2.52	0.792

Rating are based on a 5-point likert scale where 1 = strongly disagree and 5 = strongly agree

P-value: Significance of difference between class one, class two and class three: $*P < 0.05$

(g) Experiences During the Group Task

Through an open ended item, the participants were requested to briefly explain the best and worst experiences they had during the discussion period. The results from best experiences were coded into seven items which are summarized in Table 5.23 and those for worst experiences were also coded into seven items which are summarized in Table 5.24.

Table 5.23: Responses on best experiences

Benefits	Total Frequency	percentage	Class one	Class two	Class three
Learning from peers	27	30%	9	8	10
Understanding the concept	11	12%	5	3	3
Enjoyed the online discussion topic	10	11%	3	3	4
Free to critiqueize others work and offer alternatives	9	10%	2	3	4
Online learning resources and platform	9	10%	1	6	2
Social Interaction and exchange of ideas online	8	9%	2	3	3
Learning new concepts which I had not understand	3	3%	1	1	1

Table 5.24: Responses on worst experiences

Challenges	Total Frequency	percentage	Class one	Class two	Class three
Slow access to the site/slow internet	32	36%	9	14	9
Limited time	8	9%	3	3	2
Slow response from peers	7	8%	3	2	2
Difficulty Questions	5	6%	2	1	2
Lack of co-operation from group members	5	6%	1	1	3
Not reaching to an agreement in the discussion	3	3%	0	1	2
Out of topic posts	2	2%	1	1	0
Others	5	5%	1	2	2

Basically majority of the students reported that learning from peers was a good experience (27%) and it helped them understand the concepts studied (11%).

Table 5.25 presents the common themes that we identified in the students' perceived benefits in an online collaborative learning environment. For each theme, the table gives a few illustrative comments made by students.

Table 5.25: Benefits which students cited for participation in online collaborative group work

Theme	Cited example
Learning from other Peers	*It was fantastic moment since i was able to learn a lot from my peers who are doing the same course as me since people who could not contribute on face to face discussion group may be due to lack of confidence and may be didn't know how to express themselves in front of people contributed and it was just surprising to see how they had good ideas which really helped a lot during discussion.having the lecturers summarized notes online made learning easier and peaceful* *every one had there own representation of ideas as a result one was able to expand his or her way of thinking.*
Understanding the concept	*I was able to understand the topic under discussion better than when i came in.* *I experience the most effective way of learning,it built my knowledge on online skills* *It was interesting exchanging ideas online*
Learning experience was interesting	*Discussing the subject matter and giving views. The chance I got to interact with the other members in that platform was really good. It was better than face to face discussions because I could research by myself and post to the group.* *It was new, enjoyable. I got to learn about AI more than I did individually*
Free to critiqueize others work and offer alternatives	*free to critiqueize others work and offer alternatives* *I was able to voice out my answers and thoughts freely without the worrying about anything. It didn't matter whether i was right or wrong and the group members assisted me in a fine and respectable way* *Members were free to read through and agree with or point out mistakes in other people's posts where necessary thus coming to an understanding on the correct issues.*
Easy access to online learning resources	*Easy access to research material to come up with a well-founded argument.* *Digitalized and automated learning and marking respectively.*
Social Interaction and exchange of ideas online	*Interacting with my members online and being able to carry out an effective discussion* *Interacting with the group members who had randomly been chosen, thus I was able to meet new people*
Learning new concepts which I had not understand	*During the online discussion,i manage to gain a lot since we were able to openly post question and discuss the the possible answers in length unlike when we are in class.more so the discussion group minimal enough for effective discussion* *we were able to discuss areas where we were not in good in.. furthermore those i did not know i was able to know them better* *the online experience was fun and of great help because some of the things i did not understand in class I understood in the online experience*

Table 5.26 presents the common themes that we identified in the students' perceived challenges in an online collaborative learning environment. For each theme, the table gives a few illustrative comments made by students.

Table 5.26: Challenges which students cited when participating in online collaborative group work

Theme	Cited example
Slow access to the site/Slow internet	*the internet was slow leading to the delay of submission of the answers* *the network was sometimes boring and even could hinder discussion* *the system would occasionally go slow and sometimes posting things and updating would be a bit tricky*
Limited time	*less time was granted for the assignment yet we needed to carry out substantial research.* *time allocated being not so much to allow expansive discussion on the question of discussion*
Slow response from peers	*Inability to get the responses from other group members.laxity of some group members.Lack of the feedback* *delays of some members to contribute during the discussion*
Difficulty Questions	*the questions were really challenging* *experience that i encountered that seemed to be challenging was on problem solving. some members were not able to come up with a solution. although at the end of the discussion we understood he topic to some extend.*
Lack of co-operation from group members	*lack of replies and coordination from some of my group colleagues* *Intimidation by a group member when you differ with his statement even after detailing the reasons behind your differal.*
Not reaching to an agreement in the discussion	*conflicts arising due to each of the participating member viewing their idea as the masterpiece.it was part of the best but still worst experience*
Out of topic posts	*Being the group leader some members could discuss outside the topic thus difficult to come with the conclusion*

(h) Students' Suggestions on how to Improve Online Discussions

Through an open ended item, the participants were requested to briefly explain some key areas which they thought could have improved their online discussion experience. The results from 70 responses were coded into ten items which are summarized in Table 5.27. Majority felt that there was need to improve the level of participation by members (16%), provide more online discussions (9%) and increase the discussion time (7%).

On the technical side, participants felt that network connection was poor (14%), and there was a need to improve the system's interface (9%) and the site access (3%). More intervention and instructor involvement by the instructor during discussion (5%) is also necessary. A few participants felt there was need to have more group members (2%), change the group membership (2%), know group members before (1%) and have gender balance in the group (1%).

Table 5.27: Percentages of responses on students' suggestions on how to improve online discussion

Students' suggestions on how to improve online discussion	Total Frequency	percentage	Class one	Class two	Class three
Participation by all members	14	16%	4	4	6
Improve network connection	13	14%	3	6	4
More online discussion	9	10%	0	4	5
Improvements on the user interface	8	9%	0	4	4
Increase on discussion time	7	8%	5	2	0
Instructor Intervention during discussion	5	6%	1	3	1
Ability of the group leader to coordinate	3	3%	0	0	3
Having more group members	2	2%	1	0	1
Grouping of members	2	2%	1	1	0
Gender balance	1	1%	1	0	0

Table 5.28 presents the common themes that were identified in the students'suggestions for improving online collaborative learning . For each theme, the table gives a few illustrative comments made by students.

Table 5.28: Suggestions which students cited on how to improve online collaborative learning

Theme	Cited example
Participation by all members	*on the issue of participation i would have encouraged my colleagues to be more participative......in order to back up their fellow comrades reasonings...* *If everybody in the group had participated actively then it would be the best experience I have ever had.*
Improve network connection	*Make the internet connection be reliable and fast*
More online discussion	*having the discussions of the group work covering the all the unit work not only one part but involving the whole course outline* *Giving more exercises to students to do it online thus perfecting their work*
Improvements on the user interface	*copy, cut, and paste options should be included in the text editor* *Improved user interface when it comes to viewing the posts*
Increase on discussion time	*Enough time should be allocated for the discuss.Limited time may limit effectiveness of the discuss especially with net problem*
Instructor Intervention during discussion	*i would suggest that when we are holding such online discussion, lecturers of the unit should also be in a website so that we can ask them some questions and clarifications where find some difficulties.* *i think the lecturer should be part of the online class such that in case there are conflicting ideas he can guide us through to one agreement.*
Ability of the group leader to coordinate	*by selecting a group leader who can co-ordinate well the group and make the group discussion to be lively*
Having more group members	*Addition of group members for a more robust and diversified discussion and prior experience with the online system*
Grouping of members	*People should choose their own group members and they should be told to be less formal and rigid to make the experience lively.*
Gender balance	*gender balance. grouping both genders together could have made the discussions more lively*

5.2 Discussion of Results

This section examines the results of the study based on the three research questions.

Research Question 1 (RQ1)

RQ1: Which group of learners amongst the intelligently grouped (class 2), randomly grouped (class three) and instructor grouped using GPA (class one) performs best in an online collaborative environment task?

To answer this research question we formulate two hypotheses:

1. Null Hypothesis: There is no significant difference in mean scores between the three classes
2. Alternative hypothesis: There is significant difference in mean score between the three classes

From the ANOVA analysis shown in Tables 5.3, 5.6, 5.8, 5.10, 5.12 and 5.14 there was no statistically significant difference between the means for all the posttest scores. Therefore, we reject our alternative hypothesis and fail to reject the null hypothesis. This means the effectiveness of intelligent grouping algorithm is equally the same as random assignment and GPA based grouping mechanisms. Therefore, the intelligent grouping algorithm was able to generate heterogeneous groups where members have diverse backgrounds including collaboration competencies, learning capabilities and social background similar to what has been proved in random assignment.

However, the method of group formation had a slight effect on the mean scores in all posttest scores. Firstly, in forum rating where the maximum score was 10.00, the GPA method (class one) had the highest mean (7.6364), followed by random assignment (class three) with a mean of 7.5600 and lastly intelligent grouping with a mean of 7.2727. This slight difference can be explained by the observed minimum score (1.00) in class two. This minimum score could have lowered down the mean of the class slightly given that the median score for class two was higher than class three. Since this forum rating was meant to assess the quality of the posts, these three classes were equally able to construct solutions which had almost the same weight regardless of the method of group formation. There was no statistically significant difference found on the mean scores, as shown in Table 5.4.

Secondly, in Rubric Analytic tool assessment where the maximum score was 10.00, the GPA method (class one) had the highest mean score (5.5758), followed by random assignment (class three) with a mean score of 5.4412 and lastly intelligent grouping with a mean score of 5.0909 even though the median score for this class was higher than that of class three. Again the same case could have happened where few students with minimum scores could have dragged down the mean. However, the minimum scores for all the three classes were 0.00 meaning there were students who neither posted or nor replied anything on the discussion forum. The Rubric Analytic tool was meant to assess how often the students were posting new posts, attaching files and making replies hence, providing scores on the participation level in the forum. The study found that the three classes were almost equally able to participate in a similar level regardless of the method of group formation. There was no statistically significant difference found on the mean scores, as shown in Table 5.6.

Thirdly, in the quiz where the maximum score was 10.00, the intelligent grouping method (class two) had the highest mean (8.2742), followed by random assignment group (class three) with a mean score of 8.0547 and lastly GPA grouping method (class one) with a mean score of 7.5074. In this quiz, class two also had the highest median score (9.00) even though the minimum score (2.75) belonged to this class. This means class two could have obtained a higher mean score than the current one if it was not for the minimum score which probably dragged down the mean score. In addition, this was proof that the intelligent grouping algorithm had the capability to produce heterogeneous groups which had diverse collaboration competence levels. Hence, it provides a collaborative learning environment which facilitates peer learning among students due to the diverse collaboration competences provided in the group. The quiz was meant to assess individual knowledge synthesis through multiple choice questions within the domain of the discussion forum. The study found that students in the three classes were almost equally able to construct knowledge during the discussion forum, synthesize it and apply it in the quiz. Performance in the quiz was above average in all the three classes with some scoring up to the maximum (10.00). There was no statistically significance difference found on the mean scores, as shown in Table 5.10.

Fourthly, in the written test where the maximum score was 20.00, the random assignment group (class three) had the highest mean score (8.3636), followed by intelligent grouping method (class two) with a mean score of 7.7419 and lastly GPA grouping method (class one) with a mean score of 7.0606. The median scores also followed the same sequence. The written test was meant to assess knowledge construction during the discussion forum in the domain area under discussion. Findings from this study indicate that knowledge construction in all the three classes was not as expected as the mean scores were below average. This was probably due to group problems which were mentioned by the participants in the posttest questionnaire and also the challenges cited by the participants. Even though the mean scores were below average, a few students had scored highly in all the three classes with the highest scoring 17.00 out of the maximum score 20.00. The slight difference in the mean score was not statistically significant, as shown in Table 5.12.

Finally, four tests were combined to give the final grade on the forum where the maximum score was 50.00. In the final score, random assignment group (class three) had the highest mean score (28.9044), followed by intelligent grouping method (class two) with a mean score of 27.4091 and lastly GPA grouping method (class one) with a mean score of 26.7941. Although class one had the highest median score the mean score was lower in this class probably because some students performed poorly pulling the mean down. The minimum score (3.75) which belonged to class two was by far below average compared with the minimum score in class three (13.00) and in class one (9.25). This final minimum score in class two was a clear indication that there was a student in class two who did not do the entire posttest. From our observation, this student has scores only in forum rating (1.00) and in the quiz (2.75), hence the student never sat for the rest of the tests. Therefore, this minimum score could have dragged down the mean score in class two on the final scores. But all the same, class two emerged second in posttest performance as observed from the final mean scores. The slight difference in the mean score was not statistically significant, as shown in Table 5.14. Therefore, the posttest results provide empirical evidence on the capability of the intelligent grouping algorithm to group students in a desirable manner, which provides learning opportunities among peers similar to those ones realized through random assignment and GPA instructor based methods.

Research Question 2 (RQ2)

RQ2: What is the association between grouping method used and group outcomes in terms of: a) students' learning experiences; b) perceived problems; c) group leadership satisfaction and; d) group task satisfaction?

Findings from the study indicate that two major problems were experienced. Firstly, individual contribution imbalance with some members contributing less than others was highly reported in all the three classes (52%). Secondly, lack of participation feedback was also reported highly in all the three classes (48%). This coincides with other studies in which the two major problems do prevail in an online collaborative learning environment (Liu et al., 2010; Roberts and McInnerney, 2007; Capdeferro and Romero, 2012; Zorko, 2009). The mean scores for yes responses were different on the two major problems in the three classes. The GPA based assignment group (class one) had more participants experiencing contribution imbalance than the other participants who were assigned groups through intelligent grouping algorithm (class two, 41%) and random assignment (class three, 50%). This could probably be explained by the fact that the GPA based method had assigned students to groups based on their academic performance such that for each group there was a student with high GPA. These students with high GPA could have dominated the discussion because they are more knowledgeable than others causing contribution imbalance. On the other hand, intelligent grouping method had the lowest number of participants experiencing this problem. This could probably be explained by the fact that this method had grouped students based on their collaboration competence level such that for each group there was at least one student who had high collaborative competence. These students could have pulled the team together and make members collaborate more evenly with minimal contribution imbalance. With regard to these differences, there was no statistically significant relationship found among the three classes in group problems and the problems experienced as per the p-values (see Table 5.16). Therefore, the study findings indicate that group problems were not associated with the group assignment method even though the percentages differed.

On group outcomes, the study tracked a number of areas, including concept learning experiences, group leadership and group task satisfaction . Firstly, on concept learning, the overall mean was above average in all items. This means students highly agreed they were able to learn a lot from the peer learning process. This peer learning process enabled them to understand the concept of problem solving in the Artificial Intelligence course much better than it could be if it was individual learning. This coincides with other studies using the constructivist approach to learning where peer learning has been reported to be more effective in helping learners to interpret, clarify and validate their understanding through constructed dialogue and negotiation with their peers than individual learning (Garrison, 1993). Furthermore, this also supports the fact that discussion forums do support e-learning by enabling learners to actively construct knowledge by formulating ideas into words that are shared with and built on through the reactions and responses of their peers in the forum (Harasim et al., 1995). Although there was a slight mean difference on the learning experiences in the three classes, according to the p-values in Table 5.17, none of the-p-values was less than 0.05 ($p < 0.05$), hence, there was no statistically significant relationship between the group formation method and the learning experience outcome. Therefore, the study found that the learning experience outcome was similar for all learners regardless of the group formation method.

Secondly, on group leadership both the leaders' perceptions on leadership role and the group members' perceptions on the leadership were identified. Group formed using GPA (class one) had their group leaders assigned using GPA, where the student with the highest GPA value in the group was being assigned the role. Intelligent grouping method group (class two) had their leaders assigned from cluster one which had the most collaborative group as per the collaboration competence level. In random group formation group (class three), the group leader assignment was done through random assignment. Regardless of the group formation and group leader assignment method, group leaders agreed that they enjoyed playing the leadership role and this motivated them to read widely. Group members also enjoyed the role played by their leaders but they acknowledge most of the group leaders were unable to summarize the group's discussion. This was an indication that some roles like summarization and making conclusion in a discussion are more difficult to be realized through a group leader. Furthermore, there was a statistically significant relationship between the group leaders summarization role and the method of group formation with p-value=0.020 ($p<0.05$). In class one where leaders were assigned using GPA the summarization role had the lowest mean (2.47) compared with the highest mean (3.10) for class three which had random assignment. Since this study was more focused on group formation rather than group leadership, we did not have concrete data to ascertain this relationship and therefore further research which has more focus on group leadership is required to investigate the relationship between group leader assignment method and the leadership roles played in a group.

Thirdly, in group task experiences, all the items were positively rated in all the three classes. Members enjoyed working in groups and more specifically on peer learning where they are able to criticize one another and reach to consensus. Group size which was four students per group was felt to be effective and most students recommended more group work in future studies with the same group membership, with a few cited need for a change in group membership to get new experiences and exposure from new members. These group task outcome experiences were felt almost similarly in all classes regardless of the group formation technique. Therefore the study did not find any statistically significant association as observed from Table 5.22 where none of the p-values was less than 0.05 ($p < 0.05$). These outcomes coincides with other studies which have found that when group work learning is shifted from teacher control to the student peer groups, it helps learners to acknowledge their dissent, disagreements and share their doubts and students become co-constructors of knowledge rather than consumers (Bruffee, 1999).

Research Question 3 (RQ3)

RQ3: What are the students' perceived benefits and challenges of online collaborative learning?

From the best experiences which were reported by the participants, students' responses confirmed that online collaborative learning has a number of benefits including: peer learning which provides a platform to freely criticize others work and offer alternatives making the learning process enjoyable, a platform for social interaction and exchange of ideas and it provides a better opportunity for understanding concepts which are difficult to learn individually. These cited benefits truly correspond to the advantages of constructivist theory of learning (Palloff and Pratt, 1999) and the observed benefits of online collaborative learning from other studies (see Table 5.25). From the worst experiences which were reported by the participants, students' responses confirmed that slow internet was a major challenge during the online collaborative learning session. This confirms our findings from the pre-study where slow internet was mentioned as a big challenge (Muuro et al., 2014b). Due to the aforementioned benefits on online collaborative learning, universities in Kenya should increase their internet bandwidth as advised by Kashorda and Waema (2014). Few students cited inadequate responses from their peers with some groups being challenged by lack of co-operation among group members, lack of consensus in the discussion and limited time for discussion. These few challenges could have resulted from group problems which were discussed earlier.

Students' suggestions on how to improve the online discussion were in line with the observed challenges and group problems. Students suggested that there was need to improve the internet speed, improve members' participation, improve forum interface, provide instructor support and increase the discussion time. Furthermore, one student suggested that it was good to have members knowing one another before discussion and also create gender balance in the group. These suggestions show that the students were keen with the online discussion and they were ready to provide views on how to make the online collaborative learning process perfect. This was echoed by the suggestion that they would like to have more online discussions in future studies.

5.3 Summary of Study Findings

The pre-study findings revealed that, Moodle was commonly used as a platform for blended e-learning and that random group assignment method in Moodle was highly applied to form groups for online group work. Use of discussion forums was also identified as preferred method by instructors in giving online group work even though most students preferred other tools such as emails, social media platforms and mobile phones when collaborating online. Some major challenges, such as lack of participation among group members and lack of feedback from instructors were reported to be setbacks to effective online collaboration in HLIs in Kenya. This coincides with other studies in other regions (Liu et al., 2010; Capdeferro and Romero, 2012). Furthermore, despite the potential advantages of collaborative learning, some instructors do not include collaborative learning activities in their online courses. Hence, some students in blended learning are not engaged in collaborative learning and they do not realize the benefits of constructivism theory of learning. As Kashorda and Waema (2014) have noted, there is need to increase internet bandwidth in HLIs in Kenya in order to avoid the challenge of slow internet connectivity as reported by all participants in this study.

The second phase of this research centered on accomplishing research question three and four. Clustering algorithms (Skmeans and EM) were used to do data mining in discussion forums in Moodle and create collaboration competence levels and the results for the two algorithms was compared. The two algorithms had almost similar distribution of students to clusters based on collaboration competence level. These collaboration competence levels defined the learners' interdependence, synthesis and dependence characteristics in collaborative learning. The clusters which were created provided data which was used to do customized group formation and a feedback platform based on collaboration competence levels. The grouping algorithm was developed based on the ranking of learners from cluster data and has been integrated in Moodle.

The third phase of this research investigated the applicability of machine learning support in group formation in a realistic online collaborative learning environment. To that end, a true experiment was carried out in order to answer research question five and six. This experiment examined whether grouping of students based on collaboration competence level does provide online collaborative learning groups which can perform similarly in group work compared with random group assignment or group formation based on GPA. From the findings, all groups had almost similar mean scores in all posttests and shared many similar group collaboration problems, experiences and outcomes. Some obstacles to effective collaboration included: slow internet, lack of participation feedback, low participation by some members and contribution imbalance. According to the observed benefits, students seem to enjoy peer learning and peer support despite the few challenges like slow internet and few members not participating actively. Thus, as suggested by the participants, instructors should embrace online collaborative learning in their courses and provide more support to groups to ensure effective participation.

From the post test analysis, it was found that there was no statistically significant difference on the mean scores for all the three classes regardless of the group formation method. Further, apart from the group leader role on summarization of group discussion work, it was found there is no statistically significant association between the group formation method and the group problems, learning experiences, group leadership and group task satisfaction. These findings suggest that the intelligent grouping algorithm tends to form collaboration groups which seem to demonstrate similar outcomes in group problems, group experiences and learning experiences when compared with random and GPA group formation methods.

5.4 Summary

Firstly, the One-Way ANOVA analysis on the pretest results showed that the randomization effect was successful in creating heterogeneous classes which were used in the experimental study. Secondly, the study findings show that our intelligent grouping algorithm is equally good in creating groups of learners who have similar group performance as the random group formation and GPA group formation methods. This was demonstrated by One-Way ANOVA analysis on the posttest scores where all the p-values on the various posttest scores did not meet the threshold ($p<0.05$) for statistically significant difference. Thirdly, group problems and outcomes are the same regardless of the group formation method. Even though students do experience group problems, online collaborative learning has a number of benefits as cited by the students. Due to the students' perceived benefits in online collaborative learning, the peer learning process should be improved as per the students' suggestions.

CHAPTER SIX

CONCLUSIONS AND FUTURE WORK

6.1 Introduction

The research presented in this thesis has addressed multiple areas in online collaborative learning environment. Firstly, literature review was carried out to identify gaps in the research domain and also to provide theoretical foundations which were required to construct the conceptual framework. A pre-study was also conducted to understand the current situation on the use of LMSs in blended learning to support collaborative learning and the various components of online collaborative learning, which students perceived as challenging in HLIs in the Kenyan context. Secondly, ML techniques were used to cluster students in a discussion forum and create collaboration competence levels. These collaboration competence levels were used to develop an intelligent grouping algorithm and a platform to provide immediate feedback in Moodle based on learner's collaboration competence level. Finally, the impact of this grouping algorithm on group work was investigated by carrying out an experimental design in an actual online learning environment.

This chapter draws conclusions of the study. Then, it highlights the research contributions and some of the limitations encountered. At the end, some recommendations for future study are presented.

6.2 Conclusions

The successful implementation of the intelligent grouping algorithm and provision of a feedback platform in Moodle suggest that the existing group formation techniques in LMSs such as Moodle can be improved through machine learning techniques. The utilization of machine learning techniques to support group formation and feedback is timely since most of the HLIs in Kenya are faced with the problem of lack of instructor support in blended e-learning. The intelligent grouping algorithm and feedback platform in Moodle requires minimum intervention by the instructors when providing instruction support on the utilization of forums. This becomes a major advantage to those instructors who have less time to provide instruction support in online collaborative learning.

Although the participants from intelligent grouping demonstrated similar performance to random assignments method, the latter method only increases the likelihood of heterogeneity in the group, but it does not guarantee that grouping is done according to learner's collaboration competence level. Thus, our grouping algorithm has the advantage of guaranteeing heterogeneity based on learner's collaboration competence level. With the understanding that GPA group formation method involves the instructor and it may not be dynamic, instructors are more likely to adopt our intelligent grouping method as the findings show that both have similar results. Overall, it appears that the intelligent grouping algorithm provides an added advantage in supporting group formation due to its guarantee on heterogeneity, dynamism, and less instructor involvement.

6.3 Research Contributions

The purpose of this research was to investigate and experiment on ways of improving learner performance in online collaborative learning using intelligent grouping and provision of feedback in an online collaborative learning environment. In view of this, this study contributes to the body of knowledge in the following key areas:

1. **Technical:** In the software development domain, the intelligent grouping algorithm and the feedback module which have been integrated in Moodle provide a novel approach for grouping students in online collaborative learning.

Thus, this becomes a plus to the extension of Moodle and to the Software Development Community in the field of artificial intelligence . Further, the integration of data mining tools in Moodle in this study provides an opportunity to the Moodle Software Development Community on how to utilize the Moodle database in educational mining.

2. **Theoretical:** This study supplements the current literature on online collaborative learning and the use of machine learning techniques in group formation. This literature can be used in other related research to identify gaps in this domain. The research instruments which were developed in this study can guide researchers on the validation of instruments in related research. The theoretical concepts and framework which have been developed to guide system development can be utilized by other researchers to advance research in the same domain. The methodology and findings presented in this study can be used as a tool by other researchers to extend the body of knowledge in determining which group formation method or group composition should be preferred in an online collaborative learning environment. The future work highlighted in this thesis provides an opportunity to researchers to extend the body of knowledge in online collaborative learning.

3. **e-leaning pedagogy:** This research has addressed pedagogical issues such as group formation, group problems, challenges and learning experiences in online collaborative learning environment. This study informs the HLIs in Kenya the current status on online collaborative learning in terms of group collaboration problems, learning experiences, group leadership and group task satisfaction, which the current research does not address within Kenyan context. This contribution is timely since most Kenyan universities have opted to offer blended e-learning to cater for the increased demand for education online. The observed benefits and challenges create an awareness to the stakeholders of HLIs on how to improve online collaborative learning by providing ways to improve

constructivist pedagogy in e-learning. Frequent analysis of online collaborative learning through clustering techniques can provide instructors with relevant data for improving the collaborative learning process. Further, the SMS and the email tools integrated in Moodle provide a platform which instructors can use to provide immediate feedback based on learners' collaboration competence level. The positive findings on the role of group work as a learning tool from the students' perspective informs the instructors the importance of including collaborative work in instructional design. The learning experiences on the benefits and challenges on collaborative learning provides an opportunity to students with poor individual leaning skills to improve their learning through group learning. This enhances the overall quality of e-learning as well as increases the learner's confidence.

6.4 Limitations

This section outlines the main challenges which were encountered during the study and where applicable some of the measures which were taken to ensure the success of the study.

Firstly, during the experimental study there was a nation wide lecturers strike in public universities including Kenyatta University where the participants were drawn from. This derailed the experiment's design schedule as there was no learning going on. When learning resumed, instructors for the participants had to increase the number of lecture hours and discussion time to recoup the lost time. Although the sample size was 108 students, as the experiment went on, some students did not do all the posttests. The sample size reduced to 100 students and as result some groups discussed in groups of three instead of four .

Secondly, the idea of using posttest scores as Continuous Assessment Test (CAT) was applied because students tend to be more serious when doing examinations, and therefore it was assumed they will take the group task more serious if it was part of semester exam in the course. However, this could have added some pressure to all students making them read widely and respond well to the tests.

Finally, there are different group formation techniques but this study considered only three due to resources and time requirement. Other group formation techniques like self-selection were not used in the comparison groups even though they also highly used by instructors.

6.5 Future Work

Further research should be carried out to investigate instructors' level of awareness, utilization and perceived challenges of online collaborative learning tools which are available in e-learning platforms. In addition, further research should explore how online collaborative learning can be made more effective by examining the role of instructor in supporting group work, instructors' perceptions on group work and instructors level of experience in conducting collaborative learning. This could also shed more light on how to improve the quality of online collaborative learning in HLIs in Kenya.

Similar future studies should adopt large scale empirical approaches, within different universities or geographical regions in Kenya in order to confirm some of the findings observed here in other universities and also to be able to generalize the results to the larger population of Kenyan universities. Future studies could also consider examining the effectiveness of collaborative learning in enhancing student's learning skills and improving the level of knowledge construct in blended e-learning platforms.

Further research should also be conducted using a different course to confirm these findings. The idea of gender balance should be considered so that future studies have more female participants in order to study gender effect on group performance.

Since the study has verified the efficiency of clustering techniques in modeling student groups based on collaboration competence levels, future studies could focus on automating student's feedback process to reinforce the student level of collaboration using the SMS and email module which has been integrated in Moodle.

Furthermore, the study has only focused on clustering algorithms therefore there is a need to do more research with other machine learning algorithms which could also provide mechanisms for improving online collaborative learning. We also focused only on three attributes when clustering the learners and therefore, further research needs to be conducted to further other attributes with different clustering algorithms.

REFERENCES

Abdelaziz, M., Kamel, S. S., Karam, O., and Abdelrahman, A. (2011). Evaluation of e-learning program versus traditional lecture instruction for undergraduate nursing students in a faculty of nursing. *Teaching and Learning in Nursing*, 6(2):50–58.

Adjasi, C., Harvey, S., and Adyapong, D. (2008). Effects of exchange rate volatility on the ghana stock exchange. *African Journal of Accounting, Economics, Finance and Banking Research*, 3:28–47.

Aher, S. B. and Lobo, L. (2011). Data mining in educational system using weka. In *IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT)*, volume 3, pages 20–25, New York, USA. Foundation of Computer Science. ISBN:978-93-80864-71-13.

Alam, Z. M. and Rahman, A. M. (2012). Modeling volatility of bdt/usd exchange rate with garch models. *Journal of Economics and Finance*, 3(11).

Albidewi, I. and Tulb, R. (2014). Promise and challenge of e-learning-literature review. *European Scientific Journal*, 10(1).

Alfonseca, E., Carro, R. M., Martin, E., Ortigosa, A., and Paredes, P. (2006). The impact of learning styles on student grouping for collaborative learning: a case study. *User Modeling and User-Adapted Interaction*, 16(3–4):377–401.

Allen, M. W. (2003). *Michael Allen's Guide to e-Learning*. John Wiley and Sons Incorporated, Hoboken, New Jersey.

Anaya, A. and Boticario, J. (2009a). Clustering learners according to their collaboration. *Proceedings of the 13th International Conference on Computer Supported Cooperative Work in Design.*

Anaya, A. and Boticario, J. (2009b). A data mining approach to reveal representative collaboration indicators in open collaboration frameworks. *International Conference on Educational Data Mining.*

Anaya, A. and Boticario, J. (2009c). Reveal the collaboration in an open learning environment. *Lecture Notes in Computer Science.*

Anaya, A. and Boticario, J. (2010). Ranking learner collaboration according to their interactions. *The 1st Annual Engineering Education Conference.*

Anaya, A. and Boticario, J. (2011a). Application of machine learning techniques to analyze student interactions and improve the collaboration process. *Expert Systems with Applications on Computer Supported Cooperative Work in Design.*

Anaya, A. and Boticario, J. (2011b). Content-free collaborative learning modeling using data mining. special issue of user modeling and user-adapted interaction on data mining in education.

Andersen, T. G. and Bollerslev, T. (1998a). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economics Review*, 39(4):885–905.

Andersen, T. G. and Bollerslev, T. (1998b). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements and larger dependencies. *Journal of Finance*, 53(1):219–265.

Anderson, C. M., Martin, M. M., and Riddle, B. L. (2001). Small group relational satisfaction scale: Development, reliability and validity. *Communication studies*, 52(3):220–233.

Anderson, N. and Lin, C. (2009). Exploring technologies for building collaborative learning communities among diverse student populations. In *In Proceedings of the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education*, page 243, Paris, France.

Anderson, P. (2009). What is web 2.0? ideas, technologies, and implications for education.

Anjewierden, A., Kolloffel, B., and Hulshof, C. (2007). Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes. In *In International Workshop on Applying Data Mining in e-Learning*.

Antonakakis, N. (2007). Forecasting nominal exchange rate volatility in developing and industrialised countries. Master's thesis, Department of Economics :University of Strathclyde.

Antonakakis, N. and Darby, J. (2012). Forecasting volatility in developing countries. *MPRA Paper*.

Ardia, D. (2006). Bayesian estimation of the garch(1,1) with normal innovations. *MPRA Paper*, (No12985).

Ardia, D. (2007). Bayes garch: Bayesian estimation of the garch(1,1) with student–t innovations in r. *R package Journal*.

Ardia, D. and Hoogerheide, L. F. (2010a). Bayesian estimation of the garch(1,1) model with student–t innovations. *The R Journal*, 2(2).

Ardia, D. and Hoogerheide, L. F. (2010b). Efficient bayesian estimation and combination ofmodels–type models. *Tinbergen Institute Discussion Paper*, T1–46(4).

Art, B. and Lisa, E. (2008). Designing and delivering effective online nursing courses with evolve electronic classroom. *Journal of Computer Information Nursing*, 26(5):54–60.

Bacon, D. R., Stewart, K. A., and Anderson, E. S. (2001). Methods of assigning players to teams: A review and novel approach. *Simulation & Gaming*, 32(1):6–17.

Badrinath, S. G. and Chatterjee, S. (1988). On measuring skewness and elongation in common stock distributions: The case of the market index. *Journal of Business*, 61:451–472.

Bailie, R. T. and Bollerslev, T. (1989). The message in daily exchange rates: A conditional variance tale. *Journal of Business and Economic Statistics*, 7(3):297–305.

Barros, B. and Verdejo, M. F. (2000). Analysing student interaction processes in order to improve collaboration. the degree approach. *International Journal of Artificial Intelligence in Education*, 11(3):221–241.

Barrows, H. S. (1994). Practice-based learning: Problem-based learning applied to medical education. Southern Illinois University, School of Medicine, Springfield, IL.

Beg, R. A. and Anwar, S. (2012). A case study of the u.k. pound/u.s. dollar exchange rate returns. *North American Journal of Economics and Finance*, 23:165–184.

Benbunan-Fich, R., Hiltz, S. R., and Harasim, L. (2005). The online interaction learning model: An integrated theoretical framework for learning networks. in s. r. hiltz & r. goldman (eds.), learning together online: Research together online. In *Research on asynchronous learning networks*.

Benson, A. (2002). Using online learning to meet workforce demand: A case study of stakeholder influence. *Quarterly Review of Distance Education*, 3(4):443–452.

Bloom, B. S. (1956). *The Taxonomy of Educational Objectives: Classification of Educational Goals Handbook 1: The Cognitive Domain*. 1. McKay Press, New York.

Blowers, P. (2003). Using student skill self-assessments to get balanced groups for group projects. *College Teaching*, 50(3):106–110.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327.

Bollerslev, T. (1987). A conditionally heteroscedastic time series model for speculative prices and rates of returns. *The Review of Economics and Statistics*, 69(3):542–547.

Bollerslev, T., Chou, T. Y., and Kroner, K. F. (1992). Arch modelling in finance: A selective review of the theory and empirical evidence. *Journal of Econometrics*, 52:5–59.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172.

Bonk, C. J. and Dennen, V. (1999). Teaching on the web: With a little help from my pedagogical friends. *Journal of Computing in Higher Education*, 11(1):3–28.

Bonk, C. J. and Kim, K. A. (1998). Extending sociocultural to adult learning. in m. c. smith & t. pourchot (eds.). In *Adult Learning and Development: Perspectives from Educational Psychology*.

Bonk, C. J., Malikowski, S., Angeli, C., and Supplee, L. (1998). Holy cow: Scaffolding case-based "conferencing on the web" with preservice teachers. In *In American Educational Research Association (AERA) annual convention*, San Diego, CA.

Bourgerol, P. and Picard, N. (1992). Stationarity of garch processes and some non-negative time series. *Journal of Econometrics*, 52:115–127.

Brandon, D. P. and Hollingshead, A. B. (1999). Collaborative learning and computer-supported groups. *Communication Education*, 48(2):109–126.

Brindley, J., Blaschke, L., and Walti, C. (2009). Creating effective collaborative learning groups in an online environment. *The International Review of Research in Open and distance Learning*. Available at: http://www.irrodl.org/index.php/irrodl/article/viewArticle/675.

Brown, J. S. and Adler, R. P. (2008). Minds on fire: Open education, the long tail and learning 2.0. *EDUCAUSE Review*, 43(1):16–32.

Bruffee, K. A. (1999). *Collaborative Learning: Higher Education, Interdependence,* *and the Authority of Knowledge*. The John Hopkins University Press, Baltimore, MD, 2 edition.

Bruner, J. S. (1996). *The culture of education.* Harvard University Press.

Bunnett, N. and Dunne, E. (1991). The nature and quality of talk in co-operative classroom groups. *Learning and Instruction*, 1(1):103–118.

Capdeferro, N. and Romero, M. (2012). Are online learners frustrated with collaborative learning experiences? *The International review of research in open and distance learning*, 13(2):26–44. Retrieved on January 5th 2014, from: http://www.irrodl.org/index.php/irrodl/article/view/1127.

Carmona, C., Castillo, G., and Millan, E. (2007). Discovering student preferences in e-learning. In *In Proceedings of the international workshop on applying data mining in e-learning*, pages 23–33.

Chapman, K. J., Meuter, M., Toy, D., and Wright, L. (2006). Can't we pick our own groups? the influence of group selection method on group dynamics and outcomes. *Journal of Management Education*, 30(4):557–569.

Choy, M. (2002). The developments of debt mmarket in peru. *Bank for International Settlement (BIS)*, page p11.

Clark, R. (2002). Six principles of effective e-learning: What works and why. *The e-Learning Developer's Journal*, pages 1–10.

Connerley, M. L. and Mael, F. A. (2001). The importance and invasiveness of student team selection criteria. *Journal of Management Education*, 25(5):471–494.

Conrad, D. (2002). Deep in the hearts of learners: Insights into the nature of online community. *International Journal of e-Learning & Distance Education*, 17(1):1–19.

Corich, K. S. and Hunt, L. (2004). Assessing discussion forum participation: In search of quality. *Journal of Instructional Technology and Distance Learning*. Retrieved from: http://itdl.org/journal/Dec_04/article01.htm.

Craig, E. (2007). Changing paradigms: Managed learning environments and web 2.0. *Campus-Wide Information Systems*, 24(3):152–161.

Cunningham, S., Hebblewhite, R., Grout, V., and Oram, D. (2014). E-learning: The departure of the human factor in education. Retrieved on 19th September 2014.

de Faria, E. S. J., Adan-Coello, J. M., and Yamanaka, K. (2006). Forming groups for collaborative learning in introductory computer programming courses based on students' programming styles: an empirical study. In *In Frontiers in Education Conference, 36th Annual, IEEE*, pages 6–11.

Deibel, K. (2005). Team formation methods for increasing interaction during in-class group work. In *ACM SIGCSE Bulletin*, volume 37, pages 291–295, Caparica, Portugal. ACM. In: Annual Joint Conference Integrating Technology into Computer Science Education. Proceedings of the 10th annual SIGCSE Conference on Innovation and Technology in Computer Science Education.

Derek, L. (2013). Time series modelling of monthly wti crude oil returns. Master's thesis, Lincoln College: University of Oxford.

Deschamps, J. (2006). A flexible prior distribution for markov switching aautoregressive with stustudent errors. *Journal of Econometrics*, 133(1):153–190.

Devine, D. J. (2002). A review and integration of classification systems relevant to teams in organizations. *Group Dynamics: Theory, Research, and Practice*, 6(4):291–310.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263.

Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collaborative-learning: Cognitive and Computational Approaches*, pages 1–19.

Dillenbourg, P. and Schneider, D. (1995). Collaborative learning and the internet. In *International Conference on Computer Assisted*

Instruction '95. Retrieved March 12, 2011, from the web site: http://tecfa.unige.ch/tecfa/research/CMC/colla/iccai95_1.html.

Dimopoulos, I., Petropoulou, O., Boloudakis, M., and Retalis, S. (2013). "using learning analytics in moodle for assessing students' performance.". In *2nd Moodle Research Conference*, pages 40–46, Sousse, Tunisia. MOODLE.

Ding, Z., Engel, R. F., and Granger, C. W. J. (1993). Stock memory property of stock returns and a new model. *Journal of Empirical Finance*, 1:83–106.

Dringus, L. P. and Cohen, M. S. (2005). An adaptable usability heuristic checklist for online courses. In *Frontiers in Education, Proceedings 35th Annual Conference*, pages T2H–6. IEEE.

Duffy, P. and Bruns, A. (2006). The use of blogs, wikis and rss in education: A conversation of possibilities. In *Online Learning and Teaching Conference*, pages 31–38, Brisbane. Retrieved from http://eprints.qut.edu.au/5398/.

Ehlers, U. (2008). A new pathway for e-learning: From distribution to collaboration and competence in e-learning. *AACE Journal*, 16(2):187–202.

Ellis, R. (2004). Down with boring e-learning! interview with e-learning guru dr. michael w. allen. Retrieved from. http://www. astd. org/LC/2004/0704_allen. htm.

Engel, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engelhart, M. (1972). *Methods of educational research.* Rand McNally, Chicago.

Esichaikul, V., Lamnoi, S., and Bechter, C. (2011). Student modelling in adaptive e-learning systems. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 3(3):324–355.

Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business*, 381(1):34–105.

Finegold, A. and Cooke, L. (2006). Exploring the attitudes, experiences and dynamics of interaction in online groups. *The Internet and Higher Education*.

Forsyth, D. (2009). *Group dynamics*. Wadsworth Cengage Learning, Belmont, CA.

Fraenkel, J. R. and Wallen, N. E. (2000). *How to design and evaluate research in education*. McGraw-Hill.

Fraenkel, J. R., Wallen, N. E., and Hyun, H. H. (2012). *How to design and evaluate research in education*. McGraw-Hill Higher Education, New York, N.Y.

Francescato, D., Mebane, M., Porcelli, R., Attanasio, C., and Pulino, M. (2007). Developing professional skills and social capital through computer supported collaborative learning in university contexts. *International Journal of Human-Computer Studies*, 65(2):140–152.

Francq, C. and Zakoian, J. M. (2004). Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637.

Francq, C. and Zakoian, J. M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley and Sons Inc., 1st edition.

Freeman, M. (1997). Flexibility in access, interaction and assessment: The casefor web-based teaching programs. *Australian Journal of Educational Technology*, 13(1):23–39.

Gaillet, L. L. (1994). An historical perspective on collaborative learning. *Journal of Advanced Composition*, 14(1):93–110.

Gallini, J. K. and Barron, D. (2002). Participants' perceptions of web-infused environments: A survey of teaching beliefs, learning approaches, and communication. *Journal of Research on Technology in Education*, 34(2):139–156.

Gama, J. and Gaber, M. M. (2007). *Learning from Data Streams: Processing Techniques in Sensor Networks*. Springer Verlag.

Ganzel, R. (2001). Associated learning. *Online Learning.*

Garrison, D. (1993). A cognitive constructivist view of distance education: An analysis of teaching-learning assumptions. *Distance Education.*

Gaytan, J. and McEwen, B. (2007). Effective online instructional and assessment strategies. *American Journal of Distance Education.*

Geweke, J. F. (1993). Bayesian treatment of the independent student–t linear model. *Journal of Applied Econometrics*, 8(S1):19–40.

Geweke, J. F. (2004). Getting it right: Joint distribution tests of posterior simulators. *Journal of American Statistical Association*, 99(467):799–804.

Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility on the nominal excess returns on stocks. *Journal of Finance*, 48:1779–1801.

Gong, M. (2008). Personalized e-learning system by using intelligent algorithm. In *In Knowledge Discovery and Data Mining, 2008. WKDD 2008*, pages 400–401. IEEE.

Grigoriadou, M., Papanikolaou, K. A., and Gouli, E. (2006). Investigating how to group students based on their learning styles. In *In ICALT, 2006*, pages 1139–1140.

Gunawardena, C. N., Lowe, C. A., and Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, 17(4):397–431.

Hamalainen, W., Suhonen, J., Sutinen, E., and Toivonen, H. (2004). Data mining in personalizing distance education courses. In *In World Conference on Open Learning and Distance Education*, pages 1–11, Hong Kong.

Harasim, L., Hiltz, S., Teles, L., and Turoff, M. (1995). *Learning Networks: A Field Guide to Teaching and Learning Online.* MIT Press, Cambridge, MA.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their application. *Biometrika*, 57(1):97–109.

Hawkes, M. and Romiszowski, A. (2001). Examining the reflective outcomes of asynchronous computer-mediated communication on inservice teacher development. *Journal of Technology and Teacher Education*, 9(2):285–308.

Henri, F. (1992). *Collaborative Learning Through Computer Conferencing*, chapter Computer conferencing and content analysis, pages 117–136. Springer-Verlag, Berlin, Germany.

Hickey, T., Loomis, D. G., and Mohammadi, H. (2012). Forcasting hourly electricity prices using armax-garch models: An application to miso hus. *Energy Economics*, 34:307–315.

Hiltz, R. S. (1998). Collaborative learning in asynchronous learning networks, building learning communities.

Hiltz, S. R. (1990). Evaluating the virtual classroom. In *In: Harasim, L. (Ed.), Online Education*.

Hsieh, D. A. (1989). Modeling heteroscedasticity in daily foreign exchange rates. *Journal of Business and Economic Statistics*, 7(3):301–317.

Hutchings, M., Hadfield, M., Howarth, G., and Lewarne, S. (2007). Meeting the challenges of active learning in web based case studies for sustainable development. *Innovations in Education and Teaching International*, 44(3):331–343.

Ingram, A. L. (2000). Using web server logs in evaluating instructional web sites. *Journal of Educational Technology Systems*, 28(2):137–157.

Ingram, A. L. and Hathorn, L. G. (2004). *Online collaborative learning: Theory and practice*, chapter Methods for analyzing collaboration in online communications, pages 219–220. Information Science Publishing, Hershey, PA.

Israel, J. and Aiken, R. (2007). Supporting collaborative learning with an intelligent web-based system. *International Journal of Artificial Intelligence in Education*, 17(1):3–40.

Jain, A., Murty, M., and Flynn, P. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323.

Janssen, J., Erkens, G., Kanselaar, G., and Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education*, 49(4):1037–1065.

Jaques, D. and Salmon, G. (2007). *Learning in groups: A handbook for face-to-face and online environments*. Routledge, UK, USA, and Canada, 4th edition.

Johansen, R. (1988). *Groupware: Computer support for business teams*. The Free Press, New York.

Johnson, D. W. and Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Interaction Book Company, Edina, MN.

Johnson, D. W., Johnson, R. T., and Smith, K. A. (1998). Cooperative learning returns to college. *Change*, 30(4):26–35.

Johnson, D. W., J. R. T. H. E. J. . R. P. (1984). *Circles of learning: Cooperation in the classroom*. Alexandria, VA: Assoc for Supervision and Curriculum Development.

Jonassen, D. (1997). Instructional design models for well-structured and illstructured problem-solving learning outcomes. *Educational Technology: Research and Development*, 45(1):65–95.

Jonassen, D., Davidson, M., Collins, M., Campbell, J., and Banaan-Haag, B. (1995). Constructivism and computer mediated communication in distance education. *American Journal of Distance Education*, 9(2):7–26.

Jonassen, D. H. and Kwon, H. G. (2001). Communication patterns in computer mediated versus face-to-face group problem solving. *Educational Technology Research and Development*, 49(1):35–51.

Kamal, Y., Ul-Haq, H., and Khan, M. M. (2012). Modelling exchange rate volatility using generalized autoregressive conditional heteroscedasticity(garch) type models: Evidence of pakistan. *African Journal of Business Management*, 6(8):2830–2838.

Kashorda, M. and Waema, T. (2014). E-readiness survey of kenyan universities (2013). Report, Kenya Education Network, Nairobi. Retrieved on 20th August 2014, from: http://ereadiness.kenet.or.ke:8080/ereadiness/2013/E-readiness

Kaye, A. (1992). Learning together apart. in a. r. kaye. *Collaborative Learning Through Computer Conferencing*.

Khan, B. (2001). *Web-based training*. Educational Technology publications, Englewood Cliffs, NJ.

Kim, K. J., Liu, S., and Bonk, C. J. (2005). Online mba students' perceptions of online learning: Benefits, challenges, and suggestions. *The Internet and Higher Education*.

Kimball, L. (2001). *The Digital University: Reinventing the Academy*, chapter Managing distance learning: New challenges for faculty, pages 25–38. Springer Verlag, Berlin, Germany.

Koo, A.-C. (2008). Factors affecting teachers' perceived readiness for online collaborative learning: A case study in malaysia. *Educational Technology and Society*, 11(1):266–278.

Koop, G. (2003). *Bayesian Econometrics*. Number ISBN 0470845678. Wiley-Interscience, London, UK, 1st edition.

Koppenhaver, G. D. and Shrader, C. B. (2003). Structuring the classroom for performance: Cooperative learning with instructor assigned teams. *Decision sciences Journal of innovative education*, 1(1):1–21.

Koschmann, T. (1996). Paradigm shifts and instructional technology: An introduction. in t. koschmann. *Theory and Practice of an Emerging Paradigm.*

L. Valetts, B. N. and Gesa, R. (2008). Modelling collaborative competence level using machine learning techniques. *In e-Learning*, pages 56–60.

Laffey, J., Tupper, T., Musser, D., and Wedman, J. (1998). A computer-mediated support system for project-based learning. *Educational Technology Research and Development.*

Laster, S. J. (2003). Creating a blended mba program. *Syllabus*, 16(7):33–35.

Li, X. and Chang, S. K. (2005). A personalized e-learning system based on user profile constructed using information fusion. pages 109–114.

Liu, S., Joy, M., and Griffiths, N. (2009). igls: intelligent grouping for online collaborative learning. In *In Advanced Learning Technologies, 2009. ICALT 2009, Ninth IEEE International Conference*, pages 364–368. IEEE.

Liu, S., Joy, M., and Griffiths, N. (2010). Students' perceptions of the factors leading to unsuccessful group collaboration. In *In Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference*, pages 364–368. IEEE.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrica*, 65:297–303.

Longmore, R. and Robinson, W. (2004). Modelling and forecasting exchange rate dynamics: An application of asymmetric volatility models. Bank of Jamaica Working Paper WP2004/03.

Lu, F., Li, X., Liu, Q., Yang, Z., Tan, G., and He, T. (2007). Research on personalized e-learning system using fuzzy set based clustering algorithm. In *In Computational Science-ICCS 2007*, pages 587–590. Springer, Berlin, Heidelberg.

Maana, I., Mwita, P. N., and Odhiambo, R. (2010). Modelling the volatility of exmarket rates in the kenya market. *African Journal of Business Management*, 4(7):1401–1408.

Mahenthiran, S. and Rouse, P. (2000). The impact of group selection on student performance and satisfaction. *International Journal of Educational Management*, 14(6):255–265.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(3):394–419.

Mantyla, K. (2001). *Blended e-learning: The power is in the mix.* American Society for Training & Development, Alexandria, VA.

Mason, R. (1992). Evaluation methodologies for computer conferencing applications. *Collaborative Learning Through Computer Conferencing.*

Mattar, J. A. (2010). Constructivism and connectivism in education technology: Active, situated, authentic, experiential, and anchored learning. *Technology*, pages 1–16. Available at: http://www.joaomattar.com/Constructivism.

McConnell, D. (2000). *Implementing Computer Supported Cooperative Learning*. Kogan, London, 2 edition.

McCurdy, T. and Morgan, I. (1988). Testing the martingale hypothesis in deutsche mark futures with models specifying the form of heteroscedasticity. *Journal of Applied Econometrics*, 3:187–202.

McLaren, B., Scheuer, O., and Miksatko, J. (2010). Supporting collaborative learning and e-discussions using artificial intelligence techniques. *International Journal of Artificial Intelligence in Education (IJAIED).*

McLaren, B. M., Scheuer, O., Laat, M. D., Hever, R., Groot, R. D., and Rose, C. P. (2007). Using machine learning techniques to analyze and support mediation of student e-discussions. *Frontiers in Artificial Intelligence and Applications*, 158:331.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking arma time series models using squared residual autocorrelations. *Journal of Time Series Analysis*, 4:269–273.

Messeguer, R., Medina, E., Royo, D., Navarro, L., and Juarez, J. P. (2010). Group prediction in collaborative learning. In *In Intelligent Environments (IE), Sixth International Conference*, pages 350–355. IEEE.

Michaelsen, L. and Black, R. (1994). Building learning teams: The key to harnessing the power of small groups in higher education. collaborative learning. *A sourcebook for higher education*, 2:65–81.

Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Sage, Thousand Oaks, CA.

Mitchell, T. (1997). *Machine Learning.* McGraw Hill.

Mohammadi, H. and Su, L. (2010). International evidence on crude oil price ddynamic: Application of arima-garch models. *Energy Ec*, 32:1001–1008.

Moller, L. (1998). Designing communities of learners for asynchronous distance education. *Educational Technology Research and Development.*

Moore, J. L., Dickson-Deane, C., and Galyen, K. (2011). e-learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, 14(2):129–135.

Muehlenbrock, M. (2006). Learning group formation based on learner profile and context. *International Journal on E-learning*, 5(1):19–24.

Muller, T. E. (1989). Assigning students to groups for class projects: An exploratory test of two methods. *Decision Sciences*, 20(3):623–634.

Muuro, E., Wagacha, P., and R.Oboko (2014a). *Models for Improving and Optimizing Online and Blended Learning in Higher Education*, chapter 11, pages 204–219. IGI Global.

Muuro, E., Wagacha, P., R.Oboko, and Kihoro, J. M. (2014b). Students' perceived challenges in an online collaborative learning environment: A case of higher learning institutions in nairobi, kenya. *The International Review of Research in Open and Distance Learning (IRRODL)*, 15. To be published in the next issue.

Nakatsuma, T. (1998). A markov-chain sampling algorithm for garch models. *Studies in Nonlinear Dynamics and Econometrics*, 3(2):107–117.

Nelson, D. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, 59:347–370.

Newey, W. K. and West, K. D. (1987). A simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Nichols, M. (2003). A theory of elearning. *Educational Technology & Society*, 6(2):1–10.

North, A. C., Linley, P., and Hargreaves, D. (2000). Social loafing in a co-operative classroom task. *Educational Psychology*, 20(4):389–392. Retrieved on 30th July 2014 from http://www.pgce.soton.ac.uk/IT/Research/Papers/Northetal2000SocialLoafing.pdf.

Nunamaker, J. F., Chen, J. M., and Purdin, T. D. M. (1991). Systems development in information systems research. *Journal of Management Information Systems*, 7(3):89–106.

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill, New York, 2 edition.

Nyerere, J. A., Gravenir, F. Q., and Mse, G. S. (2012). Delivery of open, distance and e-learning in kenya. *The International Review of Research in Open and Distance Learning*, 13(3):185–205.

Oblinger, D. G. and Oblinger, J. L. (2005). Educating the net generation. *EDUCAUSE*. Retrieved from. http://net.educause.edu/ir/library/pdf/pub7101.pdf.

Oboko, R. and Wagacha, P. W. (2012). Using adaptive link hiding to provide learners with additional learning materials in a web-based system. *Journal of the Research Center for Educational Technology*, 8(1):11–25.

Olapiriyakul, K. and Scher, J. M. (2006). A guide to establishing hybrid learning courses: Employing information technology to create a new learning experience and a case study. *The Internet and Higher Education*, 9(4):287–301.

Olguin, C. J. M., Delgardo, A. L. N., and Ricarte, I. L. M. R. (2000). An agent infrastructure to set collaborative environments. *Educational Technology & Society*, 3(3):65–73.

Olowe, R. A. (2009). Modelling naira/dollar exchange rate volatility: Application of garch and asymmetric models. *International Journal of Busines Research Papers*, 5(3):378–398.

Oyelade, O. J., Oladipupo, O., and Obagbuwa, I. C. (2010). Application of k means clustering algorithm for prediction of students academic performance. *International Journal of Computer Science and Information Security*. Retrieved from, http://arxiv.org/ftp/arxiv/papers/1002/1002.2425.pdf.

Palloff, R. M. and Pratt, K. (1999). *Building Learning Communities in Cyberspace: Effective Strategies for the Online Classroom*. Jossey-Bass, San Francisco, CA.

Palloff, R. M. and Pratt, K. (2004). *Collaborating Online: Learning Together in Community*. Jossey-Bass.

Palloff, R. M. and Pratt, K. (2005). *Collaborating online: Learning together in community*. Jossey-Bass, San Francisco, CA.

Panitz, T. (1997). Collaborative versus cooperative learning: A comparison of two concepts which will help us understand the underlying nature of interactive learning. Retrieved on July 24, 2003 from http://home.capecod.net/ tpanitz/tedsarticles/coopdefinition.htm.

Petraglia, J. (1998). *Reality by Design: The Rhetoric and Technology of Authenticity in Education.* Lawrence Erlbaum Assoc., Mahwah, NJ:.

Piaget, J. (1929). *The Child's Conception of the World.* Brace Jovanovich, Harcourt: New York.

Qing, L. (2002). Development of the collaborative learning measure in cmc. *Journal of Educational Technology Systems*, 30(1):19–41.

Rachev, S. T., Menn, C., and Fabozzi, J. M. (2005). *Fat Tailed and Skewed Asset Distributions.* Wiley, New York.

Ramzan, S., Ramzan, S., and Zahid, F. M. (2012). Modeling and forecasting exchange rate dynamics in pakistan using arch family of models. *Electronic Journal of Applied Statistical Analysis*, 5(1):15–29.

Republic of Kenya, GoK (2005). *A policy framework for education, training and research: Meeting the challenges of education, training and research in Kenya in the 21st century.* Government Printers, Nairobi.

Rice, W. (2011). *Moodle 2.0 E-Learning Course Development.* Packt Publishing Ltd.

Roberts, T. S., editor (2004). *Online collaborative learning: Theory and practice.* IGI Global.

Roberts, T. S. and McInnerney, J. M. (2007). Seven problems of online group learning (and their solutions). *Journal of Educational Technology and Society*, 10(4):257–268.

Romero, C., Ventura, S., and Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1):368–384.

Ronteltap, F. and Eurelings, A. (2002). Activity and interaction of students in an electronic learning environment for problem-based learning. *Distance Education*, 23(1):11–22.

Roschelle, J. and Teasley, S. (1995). *Computer Supported Collaborative Learning*, chapter The construction of shared knowledge in collaborative problem-solving, pages 69–97. Springer-Verlag, Heidelberg, Germany.

Ryan, S. K. and Worthington, A. C. (2004). Market, interest rates and foreign exchange rate risk in australian banking. a garch-m approach. *International Journal of Applied Business and Economic Research*, 2(4):81–103.

Sage, S. M. (2000). The learning and teaching experiences in an online problem-based learning course. New Orleans, Louisiana, USA. Paper presented at the American Education Research Association.

Schellenberg, J. A. (1959). Group size as a factor in success of academic discussion groups. *The Journal of Educational Sociology*, 33(2):73–79. Retrieved on 30th July 2014, from http://www.jstor.org/stable/2264459.

Schneckenberg, D., Ehlers, U., and Adelsberger, H. (2010). Web 2.0 and competence-oriented design of learning-potentials and implications for higher education. *British Journal of Educational Technology*.

Scott, J. (2001). *Social Network Analysis : a handbook*. Sage, London, 2 edition.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2):4–13.

Sharma, N., Bajpai, A., and Litoriya, M. (2012). Comparison the various clustering algorithms of weka tools. *International Journal of Emerging Technology and Advanced Engineering*, 2(5). Retrieved from; http://www.ijetae.com/files/Volume2Issue5/IJETAE_0512_13.pdf.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learning*, 2(1):3–10.

Silverman, D. (1993). *Interpreting Qualitative Data*. Sage Publications, Thousand Oaks, CA.

Singh, H. K. (2005). Learner satisfaction in a collaborative online learning environment. Available at : http://asiapacific-odl.oum.edu.my/C33/F239.pdf.

Singhal, A. and Kalra, B. M. (2012). A literature review of online discussion forum in e-learning scenario. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(4):704.

Slavin, R. E. (1996). Research on cooperative learning and achievement. *What we know, what we need to know. Contemporary educational psychology*, 21:43–69.

Smith, W. (1994). Computer-mediated communication: An experimental study. *Journalism Educator*.

Soller, A., Martínez, A., Jermann, P., and Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education*.

Song, L., Singleton, E. S., Hill, J. R., and Koh, M. H. (2004). Improving online learning: Student perceptions of useful and challenging characteristics. *The internet and higher education*, 7(1):59–70.

Spacco, J., Winters, T., and Payne, T. (2006). Inferring use cases from unit testing. In *In AAAI Workshop on Educational Data Mining*, pages 1–7, New York. AAAI.

Strong, J. T. and Anderson, R. E. (1990). Free-riding in group projects: Control mechanisms and preliminary data. *Journal of Marketing Education*, 12(2):61–67.

Surjono, H. (2014). The evaluation of a moodle based adaptive e-learning system. *International Journal of Information & Education Technology*, 4(1).

Talavera, L. and Gaudioso, E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In *In Workshop on Artificial Intelligence in CSCL*, pages 17–23, Valencia, Spain.

Tang, T. and McCalla, G. (2005). Smart recommendation for an evolving e-learning system. *International Journal on E-Learning*, 4(1):105–129.

Taylor, S.

Taylor, V. (2004). *Computer supported collaborative learning in higher education*, chapter Online group projects: Preparing the instructors to prepare the students.

Tinzmann, M. B., Jones, B. F., Fennimore, T. F., Bakker, J., Fine, C., and Pierce, J. (1990). What is the collaborative classroom? Retrieved on July 24, 2003 from: http://www.ncrel.org/sdrs/areas/rpl_esys/collab.htm.

Triacca, L., Bolchini, D., Botturi, L., and Inversini, A. (2004). Mile: Systematic usability evaluation for e-learning web applications. *AACE Journal*, 12(4).

Tudge, J. (1990). Vygotsky, the zpd, and peer collaboration: Implications for classroom practice. In *In Moll, L. (Ed.), Vygotsky and Education: Instructional Implications and Applications of Sociohistorical Psychology*.

Vee, G. N. C., Gunpot, P. N., and Sookia, N. (2011). Forecasting volatility of usd/mur exchange rate using a garch (1,1) model with ged and student's-t errors. *University of Mauritius Research Journal*, 17.

Vrostos, I. D., Dellaportas, P., and Politis, D. N. (2000). Full bayesian inference for garch and egarch models. *Journal of Business and Economic Statistics*, 18(2):187–198.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, UK: Harvard University Press.

Wagner, E. D. (2001). *Web-based Training*, chapter Emerging learning trends and the world wide web, pages 33–50. Educational Technology Publications, Englewood Cliffs, NJ.

WAMZ (2013). Macroeconomic developments and convergence report. Zonal report, West African Monetary Institute, Ghana.

Wang, H. C. and Chiu, Y. F. (2011). Assessing e-learning 2.0 system success. *Computers and Education*, 57(2):1790–1800.

Wang, M., Vogel, D., and Ran, W. (2011). Creating a performance-oriented e-learning environment: A design science approach. *Information & Management*, 48(7):260–269.

Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *The Modern Language Journal*, 81(4):470–481.

WDI (2013). World development indicators. *World Bank*.

Weinberger, A., Reiserer, M., Ertl, B., Fischer, F., and Mandl, H. (2005). Facilitating collaborative knowledge construction in computer-mediated learning environments with cooperation scripts. *Barriers and biases in computer-mediated knowledge communication*, pages 15–37.

Weiss, A. A. (1986). Asymtotic theory for arch models: Estimation and testing. *Econometric Theory*, 2(1):107–131.

WEKA (2007). Available at: http://www.cs.waikato.ac.nz/ml/weka/.

Wessner, M. and Pfister, H. R. (2001). Group formation in computer-supported collaborative learning. In *In Proceedings of the 2001 international ACM SIGGROUP conference on supporting group work, ACM*, pages 24–31.

Witten, I. H. and Eibe, F. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Diego, CA.

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Diego, CA.

Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Diego, CA.

Yengin, I., Karahoca, D., Karahoca, A., and Ozcinar, Z. (2010). Being ready for the paradigm shifts in e-learning: Where is the change happening and how to catch the change. *Procedia-Social and Behavioral Sciences*, 2(2):5762–5768.

Zorko, V. (2009). Factors affecting the way students collaborate in a wiki for english language learning. *Australasian Journal of Educational Technology.*

Zurita, G., Nussbaum, M., and Salinas, R. (2005). Dynamic grouping in collaborative learning supported by wireless handhelds. *Educational Technology & Society*, 8(3):149–161.

APPENDICES

Appendix 1: Publications

This appendix contains a list of publications which have been published from this research work.

Publications

Some of the work in this thesis has been published prior to thesis submission

E. Muuro, P. Wagacha and R.Oboko, "Models for Improving and Optimizing Online and Blended Learning in Higher Education", IGI Global (2014), 204-219.

E. Muuro, P. Wagacha, R.Oboko and J. Kihoro, "Students Perceived Challenges in an Online Collaborative Learning Environment: A Case of Higher Learning Institutions in Nairobi, Kenya", *The International Review of Research in Open and Distance Learning (IRRODL)* (2014). To be published in the next issue

Conference Proceedings

Some of the work in this thesis has been presented in the following conferences prior to thesis submission

E. Muuro, P. Wagacha and R.Oboko (2014) Intelligent Grouping Based on Collaboration Competence Level (IGCCL). 2014 International eLearning Innovations Conference & Expo: 29-31 July 2014, Safari Park Nairobi, Kenya

E. Muuro, P. Wagacha and R.Oboko (2013) Students' Perceived Challenges in an Online Collaborative Learning Environment: A Case of higher learning institutions in Nairobi, Kenya. 2013 1st International Conference of the AVU, 20-22 November, Intercontinental Hotel; Nairobi, Kenya

E. Muuro, P. Wagacha and R.Oboko (2013) Intelligent Grouping Based on Collaboration Competence Level (IGCCL). 2013 International eLearning Innovations Conference & Expo: 29-30 July 2013, Safari Park Nairobi, Kenya

Appendix 2: Letter of Request of Permit to Conduct Research in KU

This appendix contains a letter of permit to conduct research in Kenyatta University.

KENYATTA UNIVERSITY

OFFICE OF THE DEPUTY VICE-CHANCELLOR (ACADEMIC)

Tel: (+254-20) 8710901-19 Ext 57481
Fax: (+254-20) 8711380
Website: www.ku.ac.ke

P.O. Box 43844-00100
Nairobi, Kenya
E-mail:dvc-acad@ku.ac.ke

Ref. KU/DVCACAD/IRT/VOL.2/302

17th February, 2014

Mr. Maina Elizaphan Muuro
C/o Dept. of Computing & Information Technology (CIT)
Kenyatta University

Dear, Muuro,

REF: REQUEST FOR PERMISSION TO CARRY OUT RESEARCH WITH KENYATTA UNIVERSITY STUDENTS IN CIT DEPARTMENT

The above subject refers.

Your request to carry out research on "**Improving Online collaborative Learning using Machine Learning Techniques**" at Kenyatta University has been approved.

On completion of your research, you are expected to submit a hard and a soft copy of your research report/thesis to our University Library and the Institute for Research Science and Technology.

Please liaise with the Director, Institute for Research Science & Technology before commencing data collection for further guidance.

Thank you.

PROF. JOHN OKUMU
DEPUTY VICE-CHANCELLOR (ACADEMIC)

c.c. Vice-Chancellor

JO/gnm

Appendix 3: Pre-study Questionnaire

This appendix contains the questionnaire which was administered for pre-study survey.

Pre-Study Survey Questionnaire

Online collaborative learning: Students perceptions

0% [ ] 100%

elearners

*

Gender

- Female
- Male

*

What is your age bracket?

Choose one of the following answers

- 15-25 years
- 26-35 years
- 36-45 years
- 46-55 years
- 56 years and above

*

Which level of study are you currently in? Indicate the programme you have enrolled in e.g. Bsc. maths, BIT, DIT, etc on the comments side.

Choose one of the following answers

- PostGraduate
- Undergraduate
- Diploma
- Certificate
- Short Course

Please enter your comment here:

*

In which university are you currently studying or did you take an online course?

Choose one of the following answers

- Jomo Kenyatta University of Science and Technology
- Kenyatta University
- Strathmore University
- USIU
- AUSI

Rate your internet skills

Choose one of the following answers

- Moderate (I know how to access emails and browse)
- Good(I know how to access emails, browse and download materials online)
- Excellent (I know how to access emails, browse, download materials and use Social media)

How many modules/ units have you ever studied online?

Choose one of the following answers

- One module
- 2-3 modules
- 4-5 modules
- more than five modules

Why did you choose to undertake an online unit/module

Check any that apply

- It was cheaper than other modes
- Parent/sponsor insisted
- My ICt skills are well polished
- My work schedule cannot allow other modes
- My home location is not favourable for modes
- It was a university requirement
- Other(Please specify)

Which gadgets do you (did you) use to access learning materials online? Comment on the right side why you prefer to use the gadget.

Check any that apply

- My Mobile Phone and Bundles
- My Desktop/LapTop and Bundles

My IPad/NotePad and Bundles

University Desktop/LapTop WIFI/Internet

Cyber Cafe internet

My own gadgets but University WIFI/Internet

Other(Please specify) ______

Indicate how often you utilize(d) the following communication tools in an online learning enviroment:

	Very Often	Often	Sometimes	Rarely	Never
Forums					
Wikis					
Workshops					
Social Medial(Face book, Twitter, etc)					
emails					
Skype					
Video Conference					
PostCads					
Google Doc					
Chats					
Telephone					

Have you ever discussed a group activity online?

Yes No

Which group discussion/communication tool do you use most to discuss group activities?

Choose one of the following answers

- Forums
- Workshops
- Wiki
- Chats
- emails
- Video Conference
- Skype
- Podcast
- Google Doc
- Social media(Facebook, Twitter, etc.)
- Telephone
- Other(Please specify)

*

Which criteria was used to asssign you to group membership in your recent group activity

Choose one of the following answers

- I assigned myself
- Assigned by Instructor
- Default assignment in Moodle
- I dont know

*

How many members were in your group?

Choose one of the following answers

- 2- 5 members
- 5-10 members
- More than 10 members
- I dont know

*

How many group activities did you do in your unit/ module?

Only numbers may be entered in this field

*

Did your group membership change during the study of the unit/module?

- Yes
- No

*

Was there a moderator/Mentor for the discussion in your group activity

- Yes
- No

*

Were you comfortable with the team members in your group?

- Yes
- No

*

Which role did the instructor play during the discussion period in your group activity?

Check any that apply

- Encouraged learners to interact with one another
- Rewarced thoughtful contributions
- Summarized key concepts
- Moderated the discussions(policing and enforcing discussion rules and policies)
- Intervened when conflict arose
- Intervened when we were stack on an isssue
- Discouraged personal criticism
- Discouraged off topic posts
- Rated the discussion
- Provided timely feedback
- Played no role
- Other(Please specify)

*

On average how many times were you accessing the discussion posts in a week?

Choose one of the following answers

- Once in a week

- 2-3 times in a week
- 4-5 times in a wek
- More than five times in a week

*

On average how many post did you send to your group activity in a week?

Choose one of the following answers

- Only Once
- 2-3 times in a week
- 4-5 times in a wek
- More than five times in a week

*

How quickly were you responding to posts related to the discussion forum in your group activity?

Choose one of the following answers

- Never Responded
- Responded with some delay
- Immediately

*

During the discussion period with your peers in your group activity, which among the following happened?

Check any that apply

- I got a feedback every time i posted an idea
- I was informed about my participation status from time to time
- My contributions were rated by the peers
- I was informed about my individual score
- I was adviced how to improve my particiaption by the instructor
- I replied to all posts
- I challenged my peers contributions
- I read all messages/posts from my peers
- The group activity improved my understanding on the topic under discussion
- Some members of the group never contributed
- Other(Please specify)

Indicate you opinion in the following issues:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disgree
While forming group the group members should have different collaborative competences.					
It 's important to have group leader to mentor others in discussion forums.					
A group leader should be among those who are most active in the discussion.					
Timely feedback from Instructors is very useful in an online group discussion.					
I often fill very frustrated whenever a reply to my request takes too long.					
Group discussion help me improve my understanding of the subject under discussion.					

In your own opinion what would hinder you from actively participating in an online group activity?

Check any that apply

- Lack of feedback from peers
- Lack of feedback from instructor
- Differences in skill/knowledge level of group members
- Low or no participation of other group members
- Workload not shared equally
- Off-topic posts in the discussion
- Single student dominating the group discussion
- Lack of leader/mentor to guide/advice you
- Lack of time to participate
- Other(Please specify) ____________________

From what you experienced during your online group activity, indicate your opinion on the following issues:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disgree
I was satisfied with the level of contact I had with my instructor					
I was satisfied with the level of contact I had with my peers					
Replies from peers on my post/requests took too long					

Briefly describe the worst online experience you have had in an online group activity

Appendix 4: Pretest Assesment Materials

This appendix contains the pretest assesment materials which were administered before the experiment.

You are logged in as Admin User (Logout)

YOUR LOGO HERE

Home ▶ Courses ▶ Miscellaneous ▶ SCO113 C4 ▶ 27 February - 5 March ▶ Quiz one ▶ Preview

Quiz navigation

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 32 33 34 35
36 37 38 39 40

Finish attempt ...

Time left 0:56:46

Start a new preview

Navigation

- Home
  - » My home
  - Site pages
  - My profile
  - Current course
    - SCO113 C4
      - Participants
      - Reports
      - General
      - 27 February - 5 March
        - Assignment 1: AI. Agents
        - Quiz one
          - Info
          - Results
      - 6 March - 12 March
      - 13 March - 19 March
      - 20 March - 26 March
      - 27 March - 2 April
      - 3 April - 9 April
      - 10 April - 16 April
      - 17 April - 23 April
  - My courses

Settings

Quiz administration
- » Edit settings
- » Group overrides
- » User overrides

Question 1
Not complete
Not graded
Flag question
Edit question

Which is used to select the particular environment to run the agent?

Select one:
- a. Environment creator
- b. Environment Generator
- c. None of the mentioned
- d. Both a & b

Check

Question 2
Not complete
Marked out of 1.00
Flag question
Edit question

The following type of an agent has happy and unhappy states.

Select one:
- a. Simple reflex agents
- b. Model based agents
- c. Utility based agents
- d. Learning agents

Check

Question 3
Not complete
Marked out of 1.00
Flag question
Edit question

In which of the following agent does the problem generator is present?

Select one:
- a. Observing agent
- b. Learning agent
- c. Reflex agent
- d. None of the mentioned

Check

Question 4
Not complete
Marked out of 1.00
Flag question
Edit question

What is Artificial intelligence?

Select one:
- a. Playing a Game
- b. Putting your intelligence into a Computer
- c. Programming with your own intelligence
- d. Making a machine intelligent

Check

Question 5
Not complete
Marked out of 1.00
Flag question
Edit question

Which is used to provide the feedback to the learning element?

Select one:
- a. Sensor
- b. None of the mentioned
- c. Actuators
- d. Critic

Check

A problem space consists of:

Select one:
- a. States
- b. none
- c. both (a) and (b)

- Edit quiz
- Preview
- » Locally assigned roles
- » Permissions
- » Check permissions
- » Filters
- » Logs
- » Backup
- » Restore
- Question bank

Course administration

Switch role to...

My profile settings

Site administration

Search

Question 6
Not complete
Not graded
Flag question
Edit question

d. Operators

Check

Question 7
Not complete
Marked out of 1.00
Flag question
Edit question

A Multi-Agent (MA) System can be described as(choose all that apply)

Select one or more:

a. A system with a coordinating intelligent behaviour among a collection of intelligent agents.

b. A group of intelligent agents processing an information set in a fixed sequence.

c. A system composed of multiple intelligent agents.

d. A collection of intelligent agents, capable of coordinating their knowledge, goals, skills and plan jointly to solve a complex problem.

Check

Question 8
Not complete
Marked out of 1.00
Flag question
Edit question

The game of Poker is a single agent.

Select one:

True

False

Check

Question 9
Not complete
Marked out of 1.00
Flag question
Edit question

What could possibly be the environment of a Satellite image Analysis System?

Select one:

a. All of the mentioned

b. Computers in space and earth

c. Image categorization techniques

d. Statistical data on image pixel intensity value and histograms

Check

Question 10
Not complete
Marked out of 1.00
Flag question
Edit question

Which is used to improve the agents performance?

Select one:

a. Observing

b. None of the mentioned

c. Perceiving

d. Learning

Check

Next

Moodle Docs for this page

You are logged in as Admin User (Logout)

SCO113 C4

You are logged in as Admin User (Logout)

YOUR LOGO HERE

Home ▶ Courses ▶ Miscellaneous ▶ SCO113 C4 ▶ 27 February - 5 March ▶ Quiz one ▶ Preview

Quiz navigation

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 32 33 34 35
36 37 38 39 40

Finish attempt ...

Time left 0:38:18

Start a new preview

Navigation

Home
- My home
  - Site pages
  - My profile
  - Current course
    - SCO113 C4
      - Participants
      - Reports
      - General
      - 27 February - 5 March
        - Assignment 1: AI. Agents
        - Quiz one
          - Info
          - Results
      - 6 March - 12 March
      - 13 March - 19 March
      - 20 March - 26 March
      - 27 March - 2 April
      - 3 April - 9 April
      - 10 April - 16 April
      - 17 April - 23 April
  - My courses

Settings

Quiz administration
- Edit settings
- Group overrides
- User overrides

Question 11
Not complete
Marked out of 1.00
Flag question
Edit question

Artificial Intelligence has its expansion in the following application. (Mark all that apply)

Select one:
- a. All of the above
- b. Diagnosis
- c. Robotics
- d. Planning and Scheduling
- e. Game Playing

Check

Question 12
Not complete
Not graded
Flag question
Edit question

What are the composition for agents in artificial intelligence?

Select one:
- a. None of the mentioned
- b. Program
- c. Both a & b
- d. Architecture

Check

Question 13
Not complete
Marked out of 1.00
Flag question
Edit question

What kind of environment is crossword puzzle?

Select one:
- a. Dynamic
- b. None of the mentioned
- c. Static
- d. Semidynamic

Check

Question 14
Not complete
Marked out of 1.00
Flag question
Edit question

What kind of behavior does the stochastic environment posses?

Select one:
- a. Local
- b. Deterministic
- c. Primary
- d. Rational

Check

Question 15
Not complete
Marked out of 1.00
Flag question
Edit question

Performance Measures are fixed for all agents. State true or false

Select one:
- True
- False

Check

Question 16
Not complete
Not graded
Flag question
Edit question

What is the rule of simple reflex agent?

Select one:
- a. Condition-action rule
- b. Simple-action rule
- c. Both a & b
- d. None of the mentioned

- Edit quiz
- Preview
- Locally assigned roles
- Permissions
- Check permissions
- Filters
- Logs
- Backup
- Restore
- Question bank

Course administration

Switch role to...

My profile settings

Site administration

Search

Check

Question 17
Not complete
Not graded
Flag question
Edit question

Which action sequences are used to acheive the agent's goal?

Select one:
- a. Retrieve
- b. Plan
- c. Search
- d. Both a & b

Check

Question 18
Not complete
Marked out of 1.00
Flag question
Edit question

What is the expansion of PEAS in task environment?

Select one:
- a. Performance, Environment, Actuators, Sensors
- b. Peer, Environment, Actuators, Sense
- c. Perceiving, Enivornment, Actuators, Sensors
- d. None of the mentioned

Check

Question 19
Not complete
Marked out of 1.00
Flag question
Edit question

What is rational at any given time depends on? Choose several.

Select one or more:
- a. The agent's percept sequence to date
- b. The agent's prior knowledge of the environment
- c. The performance measure that defines the criterion of success
- d. The actions that the agent can perform

Check

Question 20
Not complete
Marked out of 1.00
Flag question
Edit question

An omniscient agent knows the actual outcome of its actions and can act accordingly; but omniscience is impossible in reality. Rational Agent always does the right thing; but Rationality is possible in reality. State true or false

Select one:
- True
- False

Check

Next

Moodle Docs for this page

You are logged in as Admin User (Logout)

SCO113 C4

You are logged in as Admin User (Logout)

Home ▶ Courses ▶ Miscellaneous ▶ SCO113 C4 ▶ 27 February - 5 March ▶ Quiz one ▶ Preview

Quiz navigation

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 32 33 34 35
36 37 38 39 40

Finish attempt ...

Time left 0:37:38

Start a new preview

Navigation

Home
- My home
  - Site pages
  - My profile
  - Current course
    - SCO113 C4
      - Participants
      - Reports
      - General
      - 27 February - 5 March
        - Assignment 1: AI. Agents
        - Quiz one
          - Info
          - Results
      - 6 March - 12 March
      - 13 March - 19 March
      - 20 March - 26 March
      - 27 March - 2 April
      - 3 April - 9 April
      - 10 April - 16 April
      - 17 April - 23 April
  - My courses

Settings

Quiz administration
- Edit settings
- Group overrides
- User overrides

Question 21
Not complete
Not graded
Flag question
Edit question

What is perception sequence of an agent?

Select one:
- a. A periodic inputs sets
- b. None of the mentioned
- c. Both a) and b)
- d. A complete history of everything the agent has ever perceived

Check

Question 22
Not complete
Marked out of 1.00
Flag question
Edit question

How many types of agents are there in artificial intelligence?

Select one:
- a. One
- b. Four
- c. Two
- d. Three

Check

Question 23
Not complete
Marked out of 1.00
Flag question
Edit question

Which instruments are used for perceiving and acting upon the environment?

Select one:
- a. Sensors
- b. Sensors and Actuators
- c. None of the mentioned
- d. Perceiver

Check

Question 24
Not complete
Marked out of 1.00
Flag question
Edit question

What kind of environment is strategic in artificial intelligence?

Select one:
- a. Partial
- b. Deterministic
- c. Rational
- d. Stochastic

Check

Question 25
Not complete
Marked out of 1.00
Flag question
Edit question

The task Environment of an agent consists of:(choose all that apply)

Select one or more:
- a. Sensors
- b. Environment
- c. Actuators
- d. Performance Measure

Check

Which of the following best describes Distributed Artificial Intelligence(DAI)?

Select one:
- a. Several distributed intelligent agents coordinate to solve a complex problem.
- b. Intelligent agents solve their problems individually using Artificial Intelligence.
- c. It solves problems intelligently.

- Edit quiz
- Preview
- Locally assigned roles
- Permissions
- Check permissions
- Filters
- Logs
- Backup
- Restore
- Question bank

Course administration

Switch role to...

My profile settings

Site administration

Search

Question 26
Not complete
Marked out of 1.00
Flag question
Edit question

d. Incorporating techniques of Artificial Intelligence into distributed system.

Check

Question 27
Not complete
Marked out of 1.00
Flag question
Edit question

Rational agent is the one who always does the right thing. State true or false

Select one:
- True
- False

Check

Question 28
Not complete
Marked out of 1.00
Flag question
Edit question

Categorize Crossword puzzle in Fully Observable / Partially Observable.

Select one:
- a. Partially Observable
- b. Fully Observable

Check

Question 29
Not complete
Marked out of 1.00
Flag question
Edit question

Satellite Image analysis System is (Choose the one that is not Applicable)

Select one:
- a. Partially Observable
- b. Single agent
- c. Episodic
- d. Semi-Static

Check

Question 30
Not complete
Not graded
Flag question
Edit question

Which environment is called as semidynamic?

Select one:
- a. Both a & b
- b. Environment will be changed
- c. Agent performance changes
- d. Environment does not change with the passage of time

Check

Next

Moodle Docs for this page

You are logged in as Admin User (Logout)

SCO113C4

You are logged in as Admin User (Logout)

Home ▶ Courses ▶ Miscellaneous ▶ SCO113 C4 ▶ 27 February - 5 March ▶ Quiz one ▶ Preview

Quiz navigation

1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 32 33 34 35
36 37 38 39 40

Finish attempt ...

Time left 0:37:02

Start a new preview

Navigation

Home
- My home
  - Site pages
  - My profile
  - Current course
    - SCO113 C4
      - Participants
      - Reports
      - General
      - 27 February - 5 March
        - Assignment 1: AI. Agents
        - Quiz one
          - Info
          - Results
      - 6 March - 12 March
      - 13 March - 19 March
      - 20 March - 26 March
      - 27 March - 2 April
      - 3 April - 9 April
      - 10 April - 16 April
      - 17 April - 23 April
- My courses

Settings

Quiz administration
- Edit settings
- Group overrides
- User overrides

Question 31
Not complete
Not graded
Flag question
Edit question

What is the rule of simple reflex agent?

Select one:
- a. Both a & b
- b. None of the mentioned
- c. Condition action rule.
- d. Simple-action rule

Check

Question 32
Not complete
Marked out of 1.00
Flag question
Edit question

An agent is composed of?

Select one:
- a. Agent Function
- b. Perception Sequence
- c. Architecture and Program
- d. Architecture

Check

Question 33
Not complete
Marked out of 1.00
Flag question
Edit question

Agents behavior can be best described by

Select one:
- a. Perception sequence
- b. Agent function
- c. Environment in which agent is performing
- d. Sensors and Actuators

Check

Question 34
Not complete
Marked out of 1.00
Flag question
Edit question

What is the action of task environment in artificial intelligence?

Select one:
- a. Agent
- b. Observation
- c. Solution
- d. Problem

Check

Question 35
Not complete
Marked out of 1.00
Flag question
Edit question

Which instruments are required for perceiving and acting upon the environment?

Select one:
- a. None of the above
- b. Sensors and Actuators
- c. Sensors
- d. Perceiver

Check

What kind of observing environments are present in artificial intelligence?

Select one:
- a. Both a & b
- b. Fully
- c. Learning

- Edit quiz
- Preview
- » Locally assigned roles
- » Permissions
- » Check permissions
- » Filters
- » Logs
- » Backup
- » Restore
- Question bank

Course administration

Switch role to...

My profile settings

Site administration

Search

Question 36
Not complete
Not graded
Flag question
Edit question

d. Partial

Check

Question 37
Not complete
Not graded
Flag question
Edit question

What is meant by agent's percept sequence?

Select one:

a. Complete history of perceived things

b. Used to perceive the environment

c. Both a & b

d. Complete history of actuator

Check

Question 38
Not complete
Marked out of 1.00
Flag question
Edit question

Which element in agent are used for selecting external actions?

Select one:

a. Performance

b. Learning

c. Perceive

d. Actuator

Check

Question 39
Not complete
Marked out of 1.00
Flag question
Edit question

Where is the performance measure included?

Select one:

a. Rational agent

b. Task environment

c. Actuators

d. Sensor

Check

Question 40
Not complete
Marked out of 1.00
Flag question
Edit question

An 'agent' is anything that

Select one:

a.

Perceives its environment through sensors and acting upon that environment through actuators

An 'agent' is anything that

b. Takes input from the surroundings and uses its intelligence and performs the desired operations

c. All of the mentioned

d. An embedded program controlling line following robot

Check

Next

Moodle Docs for this page

You are logged in as Admin User (Logout)

SCO113 C4

Appendix 5: Posttest Questionnaire

This appendix contains the questionnaire which was administered for post-study conducted after the experiment.

Post_Study_Questionnaire

Thanks you for accepting to participate in this study. The objective of conduction this survey is to establish students' experiences in an online group work through e-learning management system.

You are hereby requested to take time to think about the answers as per your group work activity experience in SCO 113: Foundations of Artificial Intelligence. Answer the questions as truthfully as you can and whatever is unclear, you can ask for clarification from your course instructor SCO 113. Please Note that participation to this study is Voluntary.

* Required

Part A: General Information

1. **Indicate your email address:** *

____

2. **What is your gender?** *
   gender
   Mark only one oval.

   - Male
   - Female

3. **Select your group** *
   Mark only one oval.

   - Group 1
   - Group 2
   - Group 3
   - Group 4
   - Group 5
   - Group 6
   - Group 7
   - Group 8
   - Group 9

4. **Select your class** *

Mark only one oval.

- Class 1
- Class 2
- Class 3
- Class 4

5. **Which of these internet services do you use MOST FREQUENTLY to communicate with your colleagues?** *

Mark only one oval.

- email
- SMS (text message)
- Social media
- Phone calls

6. **Did you have previous knowledge on how to use Moodle e-learning platform** *

prior knowledge

Mark only one oval.

- Yes
- No

Part B: Online Collaborative Learning Experiences

7. **I experienced technical problems while doing the online group discussion** *

Mark only one oval.

- Not at all
- Few
- I don't know
- Frequently
- Unbearable

8. **If you experienced any technical problems, please specify.**

9. **Which problems did you experience in your online group discussions assignment? Tick one or more** *

Check all that apply.

- [ ] Lack of participation feedback
- [ ] Conflict and problems in reaching consensus in the group exercise
- [ ] Individual contribution imbalance with some members contributing less than others
- [ ] Problems with negotiation skills such that it was difficult to agree on a common goal
- [ ] Lack of coordination from Group Leader
- [ ] Posting of irrelevant comments by members
- [ ] Other, specify in the next question

10. **Specify any other problems you may have experienced in online group discussions**

____

11. **To what extent would you agree or disagree with the following statements with regard to the Discussion Forum(Group Activity) you have just had. Please mark only one oval per row.** *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I found it easy to learn through the Discussion Forums in Moodle.	⬭	⬭	⬭	⬭	⬭
The participation in Group Activity improved my understanding on the topic under discussion	⬭	⬭	⬭	⬭	⬭
I learnt more about the subject matter under discussion in the group exercise than I would if I worked individually	⬭	⬭	⬭	⬭	⬭
By reading the contribution of others I had a better understanding of the problem solving concept in AI.	⬭	⬭	⬭	⬭	⬭
When group members created new post it provided me with a new perspective of the topic in discussion.	⬭	⬭	⬭	⬭	⬭
In online discussion forums I learnt more than discussions in other face to face (Lecture) methods.	⬭	⬭	⬭	⬭	⬭

12. **Were you a group leader? ***
Mark only one oval.

- Yes *Skip to question 13.*
- No *Skip to question 16.*

Skip to question 13.

13. **On the process of leading your group to what extent would you agree or disagree with the following statements . Please mark only one oval per row. ***
Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I enjoyed playing the role of a Leader in in my group					
The responsibility motivated me to read widely enabling me to lead other members in the group					
Members in my group were reluctant to contribute					

14. **How easy was it to lead your group ? Please mark only one oval. ***
Mark only one oval per row.

	Very hard	Hard	I don't know	Easy	Very easy
How easy it was to lead the group					

15. **As a Group Learder in that group activity, rate your leadership skills. Please mark only one oval. ***
Mark only one oval per row.

	Below average	Average	I don't know	Good	Very Good
My leadership Skill					

Skip to question 17.

16. **In your own opinion to what extent would you agree or disagree with the following statements . Please mark only one oval per row. ***

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I enjoyed working with my group leader					
Our group leader coordinated the group exercise well and kept the group "on-track "(kept the group focused and organized)					
Our group leader summarized the group's discussion and came up with the conclusions					
Our group leader managed conflict and differences of opinions within the group task					
Our group leader demonstrated thorough understanding of the subject content					

17. **If i was to do another group activity i will recommend a change of group members ***

Mark only one oval.

- Yes
- No

18. **Give reasons for your answer in the previous question (above). ***

19. **In regard to your experience during group work, to what extent would you agree or disagree with the following statements . Please mark only one oval per row. ***
Mark only only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
In our group activity, I was able to negotiate with my peers and reach to a consensus					
In my group activity, members were free to criticize each other contribution in a positive and constructive manner.					
I was satisfied with the level of contact I had with my peers					
Time allocated was enough to complete the group activity					
I enjoyed working with my peers in our group activity					
I would recommend for more group activities with my group members					
I think all our group members were given fair opportunity to contribute					
One or two members dominated the group exercise					
The group size was optimum for effective discussion					
I would recommend online discussion forums in future studies in my course work					

20. **Briefly describe the best online experience you had during your online group work assignment. ***

21. **Briefly describe the worst online experience you had during your online group work assignment. ***

22. **Suggest some key areas which you think could have improved your online discussion experience.** *

Powered by

Google Forms

Appendix 6: Posttest Assesment Materials

This appendix contains the posttest assesment materials which were administered after the experiment.

Post-test: Discussion questions for nine groups (one question for each group)

1. Imagine you are given two buckets (X and Y) which are not calibrated, but when the buckets are full X can hold 10 litres of kerosene and Y can hold 7 litres of kerosene. You also have a pump that can be used to fill either X or Y with kerosene and you can empty the contents of either X or Y at any time. You goal is to get exactly 1 litre of kerosene into bucket Y.
   a) Describe the state space representation for this problem and identify the initial state and goal state.
   b) Identify the rules/operators which can be used to solve this problem
   c) Apply the above rules and describe the possible solutions to this problem
   d) Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

2. Imagine you have three Cups (X, Y, Z) measuring 9 litres, 5 litres and 4 litres respectively when full, but they are not calibrated and a keg tap is available which can be used to fill them. You can fill the cups or empty them out from one to another or empty by pouring onto the ground. You goal is to serve a customer with a cup of keg measuring exactly 2 litres.

   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

3. Imagine you have four bottles (A, B, C and D) measuring 7 litres, 5 litres, 6 litres and 12 litres respectively when full but they are not calibrated. You can connect them to a water pump and fill the bottles or empty them out from one to another or empty by pouring onto the ground. Your goal is to use bottle A, B and C to obtain exactly 10 litres in Bottle D.

   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

e.

4. Imagine you have three animals (X, Y, and Z) on one side of a bridge which you wish to cross them on the other side of the bridge using a motorbike. The motorbike can only hold two items including the rider (yourself) at any one time and of course you are the only one who can ride. If the animal X is ever left alone with animal Y, X will eat it. Similarly if the Y is ever left alone with the Z, then Z will eat it. The goal is to get all the animals and yourself on the other side of the bridge safely using the motorbike.
   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

5. Three people (A, B, and C) and three animals (X, Y, Z) which are carnivorous are on one side of a river, along with a boat that can hold at most two people. For the group of people to be save you need to make sure they are never outnumbered by the carnivorous when left together. The goal is to find a way to get everyone to the other side without ever leaving a group of people in one place outnumbered by the carnivorous in that place.
   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

6. Imagine you have three items to take into the market (X, Y and Z) and you come across a river which you must cross on the way to the market. On the shore there is a boat which can only take one item at a time to the market. To have all the items save this rule must be observed item: Item Y cannot be left together with item Z and item Z cannot be left together with item X. The goal is to cross the river and have all the items taken into the market safely.
   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

7. Imagine you have been given a task to sell a product in five cities (A, B, C, D and E). To complete the task you must visit each city exactly once. There are direct roads between each pair of cities on the list. You goal is to find the shortest route possible which can take you round and visit all the cities exactly once.

   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

8. Imagine there are three guys (X, Y and Z) who want to cross a river from the left side to the right side using a boat. While you are in the boat, they behave. But as soon as you leave X with Y or X with Z on one side WITHOUT the boat, they start fighting. Being a peace loving person, you don't want that. The goal is to move them across the river with no fights. The boat can carry up to two passengers including yourself and cannot move by itself.

   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

9. Three jealous husbands (X, Y and Z) and their wives (XW, YW and ZW) need to cross a river using a single boat. At no time should any of the women be left in company with any of the men, unless her husband is present. The boat can carry up to two passengers and cannot move by itself.

   a. Describe the state space representation for this problem and identify the initial state and goal state.
   b. Identify the rules/operators which can be used to solve this problem
   c. Apply the above rules and describe the possible solutions to this problem
   d. Suggest a heuristic function which can be used to guide the search in order to obtain an optimal solution.

You are logged in as Admin User (Logout)

YOUR LOGO HERE

Home ▶ Courses ▶ Miscellaneous ▶ SCO113 C1 ▶ 27 March - 2 April ▶ Quiz 2 ▶ Preview

Quiz navigation

1 2 3 4 5 6 7 8 9 10

Finish attempt ...

Time left 0:09:00

Start a new preview

Navigation

Home

» My home

Site pages

My profile

Current course

My courses

Settings

Quiz administration

Course administration

Switch role to...

My profile settings

Site administration

Search

Question 1
Not yet answered
Marked out of 1.00
Flag question
Edit question

A solution to a problem is a path from the initial state to a goal state. Solution quality is measured by the path cost function, and an optimal solution has the highest path cost among all solutions.

Select one:

a. False

b. True

Question 2
Not yet answered
Marked out of 1.00
Flag question
Edit question

In an uninformed search technique, the search algorithm is guided by ;

Select one or more:

a. The initial state

b. Path cost

c. Heuristic function

d. Goal state

Question 3
Not yet answered
Marked out of 1.00
Flag question
Edit question

.... are means for transforming the problem from one state to another.

Select one:

a.
Search tree

b. Heuristic functions

c.
States

d. Operators/ Rules

Question 4
Not yet answered
Marked out of 1.00
Flag question
Edit question

What is state space?

Select one:

a. A space where you know the solution

b. Your definition to a problem

c. The whole problem

d. Representing your problem with variable and parameter

e. Problem you design

Question 5
Not yet answered
Marked out of 1.00
Flag question
Edit question

A problem in a search space is defined by?

Select one:

a. Goal test

b. All of these

c. Path cost

Question 6
Not yet answered
Marked out of 1.00
Flag question
Edit question

How many parts does a problem consists of?

Select one:

a. 1

b. 3

c. 2

d. 4

Question 7
Not yet answered
Marked out of 1.00
Flag question
Edit question

The major component/components for measuring the performance of problem solving are;

Select one or more:
- a. Space complexity
- b. Time complexity
- c. Optimality
- d. Completeness

Question 8
Not yet answered
Marked out of 1.00
Flag question
Edit question

.... is a set of possible permutation that can be examined by a search method in order to find a solution.

Select one:
- a. Fraction
- b. Search space
- c. None of these
- d. Formula

Question 9
Not yet answered
Marked out of 1.00
Flag question
Edit question

The best way to represent a state space for a problem is through;

Select one:
- a. Graph
- b. Predicate calculus
- c. Tree
- d. Network path

Question 10
Not yet answered
Marked out of 1.00
Flag question
Edit question

A solution path can be defined as ;

Select one:
- a. Goal state
- b. Goal test
- c. An optimal solution
- d. A path from the initial state to the goal state

Next

Moodle Docs for this page

You are logged in as Admin User (Logout)

SCO113 C1

SCO 113: C.A.T

03/04/10

Instructions:

Refer to the discussion question in your group discussion forum and answer the following questions:

1. In your discussion question what makes up a state? **[1 Marks]**

2. What guided you when formulating the state space of this problem? **[1 Marks]**

3. In order to define a solution for your discussion question, how many components for the problem did you identify, state them: **[4Marks]**

4. Identify at least six to ten possible moves which can apply to this problem. Hint for every move describe the conditions/rules, new states and description of the state. **[4 Marks]**

5. By using a search tree describe the state space for this problem **[6 Marks]**

6. From you search tree do you think this problem have more than one solution path? Justify you answer. **[2 Marks]**

7. In your discussion you were required to come up with a heuristic function. By referring to one of the heuristic function explain how it could be applied to obtain an optimal solution for this problem. Use the heuristic function to identify the optimal solution from your search tree. **[2 Marks]**

Appendix 7: Program Code Segment

This appendix contains the program code segments which were implemented for the clustering, intelligent grouping algorithm and feedback platform in Moodle.

```
//WEKA CLUSTER PHP
<?php
require_once('../config.php');
require_once('smtpmail.php');
require_once('single_message.php');

$courseid = required_param('id', PARAM_INT);
$confirm   = optional_param('confirm', 0, PARAM_BOOL);
$clusid=optional_param('clusid', 1, PARAM_INT);
$sms=optional_param('sms','',PARAM_ALPHA);
$mail=optional_param('mail','',PARAM_ALPHA);
global $DB;
$PAGE >set_url('/grouped/wekacluster.php', array('id' => $courseid));
if (!$course = $DB >get_record('course', array('id'=>$courseid))) {
    print_error('invalidcourseid');
}
// Make sure that the user has permissions to manage groups.
require_login($course);
$context        = get_context_instance(CONTEXT_COURSE, $courseid);
$systemcontext = get_context_instance(CONTEXT_SYSTEM);
require_capability('moodle/course:managegroups', $context);
$PAGE >set_pagelayout('admin');
$PAGE >set_heading($course >fullname. ': '.$strgroups);

echo $OUTPUT >header();
?>
<script>
function sendmessage(clusterid)
{
    var msg = document.getElementById('message_'+clusterid);
    if (msg.value == '') {
```

```
        alert("Empty message can't be send");
    } else {
    document.getElementById("success").innerHTML='<img src="loading.
        gif" />';
if (window.XMLHttpRequest)
  {// code for IE7+, Firefox, Chrome, Opera, Safari
  xmlhttp=new XMLHttpRequest();
  }
else
  {// code for IE6, IE5
  xmlhttp=new ActiveXObject("Microsoft.XMLHTTP");
  }
xmlhttp.onreadystatechange=function()
  {

  if (xmlhttp.readyState==4 && xmlhttp.status==200)
    {
    document.getElementById("success").innerHTML=xmlhttp.responseText
        ;
    }
  }
xmlhttp.open("GET","sendstatus.php?clusterid="+clusterid+"&msg="+msg.
   value,true);
xmlhttp.send();
    }
}
</script>
<script type="text/javascript">
function validateForm()
{
var x=document.forms["myForm"]["noofclusters"].value;

  var num_regex = /^\d+$/;
if (x==null || x=="")
  {
```

```
  alert("Number of Clusters must be filled out");
  return false;
  }

if ( !x.match(num_regex) ) {
  alert("Number of clusters must be Integer");
  return false;
}
}
</script>
<!  Form to enter no of clusters to built    >
<form name="myForm" method="post" onsubmit="return validateForm()" >
    <h2>Create Clusters</h2>
  <table>
  <tr>  <td><label>Clustering Type</label></td>
                  <td> <select  name="algtype">
                   <option>Skmeans</option>
                   <option>EM</option>
                </select></td></tr>
              <tr><td><label>Number of Clusters</label></td>

                <td><input type="text" maxlength="5"  name="
                   noofclusters"
                              value="" /></td></tr>
                      <tr>  <td></td> <td><input type="submit" name
                         ="noofclrs"
                      value="Next" /></td>          </tr>
                </table>
</form>

<!  form ends   >
<?php
                if(isset($_POST['noofclrs'])){
                $i = $_POST['noofclusters'];
```

```
//Executing java Class file
$file = 'data.csv';

if($_POST['algtype']=='EM'){
exec("java ClassesToClusters $i $file 2>&1",$output);
} else {
exec("java MyKSVM $i $file 2>&1",$output);
}

//Reading data from CSV file
                $row = 1;

//Deleting Existing Records from Temporary Folder
                $DB >delete_records("cluster_temp");
//Reading from output CSV  file which comes from weka
        if (($handle = fopen($output[0], "r")) !== FALSE) {
              while (($data = fgetcsv($handle, 1000, ";"))
                 !== FALSE) {
          $num = count($data);
          $row++;

//Storing into data object
                                  $datas = new stdClass();
                                  $datas >userID = $data[1];
                                  $datas >clusterID = $data[2];

//Inserting data into cluster_temp table
                                 $DB >insert_record('cluster_temp',
                                    $datas);
                          }// end of CSV data Read
                                  fclose($handle);
                 } //Closing CSV File

//noofusers , noofclusters
                 $noofusers  =  $DB >count_records_sql(
```

```
                                   ("SELECT count(userID) FROM {$CFG >
                                      prefix}cluster_temp");
                $noofclusters  =  $DB >count_records_sql
                                   ("SELECT count(distinct(clusterID))
                                      FROM {$CFG >prefix}cluster_temp");
          //        $noofgroups = $_POST['noofgroups'];
//Displaying the results
          echo "<p>Total Number of Students: $noofusers </p>";
           echo "<p>Total Number of clusters:$noofclusters </p> ";
            for($i=0;$i<$noofclusters;$i++){
                   $cluster[] = $DB >count_records_sql
                       ("SELECT count(clusterID) FROM {$CFG >prefix}
                          cluster_temp
                                    WHERE clusterID=$i");
                       }
               $clusterids = $DB >get_records_sql("SELECT clusterID,
                  userID
                                   from {$CFG >prefix}cluster_temp ORDER
                                      BY clusterID ASC");
                   foreach($clusterids as $clusterid){
                     $clus[] = $clusterid >clusterid;
                       }
                       $c=0;
                 foreach($cluster as $clusterno){
                 echo "<p>Number of Students in Cluster$clus[$c] :
                    $clusterno</p>";
                       $c++;
                       }

//Displaying Cluster assignments table
                  // echo "<h2>Clusters</h2>";
                   // echo "<div style='overflow:scroll;'><table>";
                    // foreach($clus as $clusno){
                   // echo "<tr><th style=\"border:1px solid gray;
                                          \">Cluster$clusno</th><form
```

```
>";
// $users = $DB >get_records_sql("SELECT ct.id ,u.username
    ,u.id
                    FROM {$CFG >prefix}user u,{$CFG >
                        prefix}cluster_temp ct
                    WHERE ct.clusterid = $clusno AND u.id
                        = ct.userid");
        // echo "<td style=\"border:1px solid gray;\">
        // <textarea rows='5' cols='20' id='message_$clusno'
              name='message'></textarea><input
        type='button' value='send' style='float:right;'
        onclick='sendmessage($clusno)'/></form></td>";
          // foreach($users as $user){
    // echo "<td style=\"border:1px solid gray;\">$user >id
        </td>";
                    // }

                    // }
       // echo "</tr></table></div><div id='success'></div
          >";

                            echo "<h2>Clusters</h2>";

            echo "<div style='overflow:scroll;'><table>";
            foreach($clus as $clusno) {
echo '<tr><th style="border:1px solid gray;">
 <a href="members_cluster.php?grd='.$clusno.'&id='.$courseid
    ."
              >Cluster '.$clusno.'</a></th><form>';
    $users = $DB >get_records_sql("SELECT ct.id ,u.username ,u.
        id
              FROM {$CFG >prefix}user u,{$CFG >prefix}
                  cluster_temp ct
              WHERE ct.clusterid = $clusno AND u.id = ct.
                  userid");
```

```
            // echo "<td style=\"border:1px solid gray;\">
           // <textarea rows='5' cols='20' id='message_$clusno'
name='message'></textarea><input type='button' value='send'
style='float:right;' onclick='sendmessage($clusno)'/></form></td>";

        $mems= array();
    foreach($users as $user){
            if ($i !=0) {
                    $mems[]=$user >username;
                } else {
                  $mems[]=$user >username."(Mentor)";
                }
                $i++;
        }
        //$userlist=implode(',',$mems);
        $x = count($mems);
                $y = '';
                for ($i=0; $i <$x; $i++ ) {
                        $x1 = ($x/2);
                        if($i % $x1 == 0 && $i !=0 && $x != 1 ) {
                                $y .= $mems[$i];
                                $y.= '</br>';
                        } else if ($x == 1)
                        $y .= $mems[$i];
                        else
                        $y .= $mems[$i] . ', ';
                }
        echo "<td style=\"border:1px solid gray;\">".$y."</td>";
        echo '<td style="border:1px solid gray;">'.$OUTPUT >
            single_button(new moodle_url('/grouped/wekacluster.php' ,
            array('clusid'=>$clusno , 'mail'=>'mail' ,'id'=>$courseid)
            ), 'Mail')."</td>";
                echo '<td style="border:1px solid gray;">'.$OUTPUT >
                    single_button(new moodle_url('/grouped/wekacluster
                    .php' , array('clusid'=>$clusno ,'sms'=>'sms', 'id
```

```
'=>$courseid)), 'sms')."</td>";
    }
                        echo "</tr></table></div><div id='
                            success'></div>";
} else {
                        $noofusers = $DB >count_records_sql
                            ("SELECT count(userID) FROM {$CFG
                            >prefix}cluster_temp");
            $noofclusters = $DB >count_records_sql("SELECT
                count(distinct(clusterID)) FROM {$CFG >prefix}
                cluster_temp");
            $noofgroups = $_POST['noofgroups'];
//Displaying the results
                  echo "<p>Total Number of Students: $noofusers
                      </p>";
                  echo "<p>Total Number of clusters:
                     $noofclusters </p> ";
                  for($i=0;$i<$noofclusters;$i++){
                        $cluster[] = $DB >count_records_sql("
                            SELECT count(clusterID) FROM {$CFG
                            >prefix}cluster_temp WHERE
                            clusterID=$i");
                  }
                  $clusterids = $DB >get_records_sql("SELECT
                     clusterID, userID from {$CFG >prefix}
                     cluster_temp ORDER BY clusterID ASC");
                    foreach($clusterids as $clusterid){
                        $clus[] = $clusterid >clusterid;
                  }
                  $c=0;
                  foreach($cluster as $clusterno){
                        echo "<p>Number of Students in
                            Cluster$clus[$c] : $clusterno</p
                            >";
                        $c++;
```

```
}
                                        echo "<h2>Clusters </
                                           h2>";

                            echo "<div style='overflow:scroll;'><
                               table>";
foreach($clus as $clusno) {
    echo '<tr><th style="border:1px solid gray;"><a href="
        members_cluster.php?grd='.$clusno.'&id='.$courseid.'"
            >Cluster '.$clusno.'</a></th><form>';
    $users = $DB->get_records_sql("SELECT ct.id,u.username,u.id
        FROM {$CFG->prefix}user u,{$CFG->prefix}cluster_temp ct
        WHERE ct.clusterid = $clusno AND u.id = ct.userid");
                            // echo "<td style=\"border:1px solid
                                gray;\">
                                  // <textarea rows='5' cols='20'
                                     id='message_$clusno' name='
                                     message'></textarea><input
                                     type='button' value='send'
                                     style='float:right;' onclick
                                     ='sendmessage($clusno)'/></
                                     form></td>";

            $mems= array();
            foreach($users as $user){
                    if ($i !=0) {
                            $mems[]=$user->username;
                    } else {
                            $mems[]=$user->username."(Mentor)";
                    }
                    $i++;
            }
            //$userlist=implode(',',$mems);
            $x = count($mems);
            $y = '';
```

```
            for ($i=0; $i <$x; $i++ ) {
                    $x1 = ($x/2);
                    if($i % $x1 == 0 && $i !=0 && $x != 1 ) {
                            $y .= $mems[$i];
                            $y.= '</br>';
                    } else if ($x == 1)
                    $y .= $mems[$i];
                    else
                    $y .= $mems[$i] . ', ';
            }
            echo "<td style=\"border:1px solid gray;\">".$y."</td
                >";

            echo '<td style="border:1px solid gray;">'.$OUTPUT >
                single_button(new moodle_url('/grouped/wekacluster
                .php' , array('clusid'=>$clusno , 'mail'=>'mail'
                ,'id'=>$courseid)), 'Mail')."</td>";
            echo '<td style="border:1px solid gray;">'.$OUTPUT >
                single_button(new moodle_url('/grouped/wekacluster
                .php' , array('clusid'=>$clusno ,'sms'=>'sms', 'id
                '=>$courseid)), 'sms')."</td>";
            }
            echo "</tr></table></div><div id='success'></div>";
}

// cluster sms and mail

if(isset($clusid) && $clusid != 1 ) {
        $results=$DB >get_records_sql("SELECT ct.id,u.email, u.id
           FROM {$CFG >prefix}user u,{$CFG >prefix}cluster_temp ct
           WHERE ct.clusterid = $clusid AND u.id = ct.userid");
        $toaddress = array();
        $numbers =array();
        foreach ($results as $result){
                $toaddress[] = $result >email;
```

```
                    $numbers[]=$result >id;
        }
        $names= $toaddress;

}

if(isset($mail) && $mail =='mail') {
        echo $OUTPUT >box_start('generalbox');
        echo '<h3>Enter Subject and  Message to send email</h3></br
            >';
        echo '<form method="post"  action="wekacluster.php?id='.
            $courseid.'&clusid='.$clusid.'">
                              <table>
                              <tr><td><lable for="subject">Subject
                                  :</lable></td><td><input type="
                                  text" name="subject"></td></tr>
                              <tr><td><lable for="message">Message
                                  :</lable></td><td><textarea rows
                                  ="4" cols="50" name="message"></
                                  textarea></td></tr>
                              <tr><td></td><td><input type="submit"
                                   value="Submit" name="button1">&
                                  nbsp <button onclick="location
                                  .reload()">Cancel</button></td></
                                  tr>
                              </table>
                    </form>';
        echo $OUTPUT >box_end();
}

if (isset($_POST['button1'])) {
 $subject = $_POST['subject'];
 $message = $_POST['message'];
 foreach ($names as $to) {
  if(smtpmailer($to, $subject, $message)) {
```

```
  echo  ''. $to .'  Mail sent successfully! <br />';
 }
 else {
  echo'sorry mail not sent to' .$to. '</br>';
 }
}

}




if(isset($sms) && $sms =='sms') {
  echo $OUTPUT > box_start('generalbox');
  echo '<h3>Enter Message to send SMS</h3></br>';
  echo '<form method="post"  action="wekacluster.php?id='.$courseid
      .'&clusid='.$clusid.'">
  <table>
  <tr><td><lable for="message">Message:</lable></td><td><textarea
      rows="4" cols="50" name="message"></textarea></td></tr>
  <tr><td></td><td><input type="submit" value="Submit" name="button2
      ">  <button onclick="location.reload()">Cancel</button
      ></td></tr>
  </table>
  </form>';
  echo $OUTPUT > box_end();
}
if (isset($_POST['button2'])) {

        if (!empty($_POST['message'])) {
                $message = $_POST['message'].'</br>';
                echo '<table border="1"><th><b>User</b></th><th><b>
                    Messagestatus</b></th>';
                for ($i=0; $i < count($numbers);$i++) {
                        message($message, $numbers[$i]);
                }
```

```
            echo "</table>";
        } else {
            echo "<h3 <h3 style='color:red;'>>You can not send
                sms with empty message.</h3>";
        }
}

echo $OUTPUT > footer();

 if(!isset($_POST['noofclrs'])){
/* vars for export */
// database record to be exported
$db_record = 'mdl_summary';

// database variables
$hostname = $CFG > dbhost;
$user = $CFG > dbuser;
$password = $CFG > dbpass;
$database = $CFG > dbname;

// Database connecten voor alle services
$conn = mysql_connect($hostname, $user, $password)
or die('Could not connect: ' . mysql_error());

mysql_select_db($database)
or die ('Could not select database ' . mysql_error());

$query = "SELECT role.userid,count(if(post.parent=0 ,post.userid,NULL
    )) AS Numberofpost,count(if(post.parent!=0 ,post.id,NULL)) AS
    Numberofreplies,
 (select round(COALESCE(avg(rate.rating),0)) from mdl_rating AS rate
     where post.id=itemid and rate.component='mod_forum' and rate.
     ratingarea='post') as avgrating
                                              FROM
                                              mdl_context AS context
```

```
INNER JOIN
    mdl_role_assignments AS
    role
ON role.contextid=context.id
    and role.roleid=5
LEFT JOIN mdl_forum_posts AS
    post
ON role.userid=post.userid

WHERE
context.instanceid=$courseid
    and context.contextlevel
    =50
group by role.userid";
$result = mysql_query( $query
    , $conn ) or die(
    mysql_error( $conn ) );


        $out ="";
  $file = fopen('data.csv','w');

        $field = mysql_num_fields($result);

// create line with field names
for($i = 0; $i < $field; $i++) {
   if($i==($field 1)){
         $out .=mysql_field_name($result,$i);
    } else {
  $out .=mysql_field_name($result,$i).',';
}
}
        $out .= "\n";
        // Add all values in the table
        while ($l = mysql_fetch_array($result)) {
```

```
for($i = 0; $i < $field; $i++) {
    if($i==($field 1)){
        $out .=$l[mysql_field_name($result ,$i)];
   } else {
 $out .=$l[mysql_field_name($result ,$i)].',';
 }
 }

        $out .="\n";
        }

        // Output to browser with appropriate mime type
        fputs($file ,$out);
        fclose($file);
        exit;
}
 ?>
```

```
//WEKA GROUP PHP
<?php
require_once('../config.php');
require_once('single_message.php');
require_once('smtpmail.php');
global $DB;
$courseid = required_param('id', PARAM_INT);
$confirm   = optional_param('confirm', 0, PARAM_BOOL);
$grd=optional_param('grd',0,PARAM_INT);
$sms=optional_param('sms','',PARAM_ALPHA);
$mail=optional_param('mail','',PARAM_ALPHA);
$PAGE >set_url('/grouped/grouping_mem.php', array('courseid' =>
    $courseid));
if (!$course = $DB >get_record('course', array('id'=>$courseid))) {
     print_error('invalidcourseid');
}
// Make sure that the user has permissions to manage groups.
```

```
require_login($course);
$context       = get_context_instance(CONTEXT_COURSE, $courseid);
$systemcontext = get_context_instance(CONTEXT_SYSTEM);
require_capability('moodle/course:managegroups', $context);
$PAGE >set_pagelayout('admin');
$PAGE >set_heading($course >fullname. ': '.$strgroups);
$gm = $DB >get_records_sql("SELECT id,name FROM mdl_groups WHERE
    courseid=$courseid");
$grname =array();

echo $OUTPUT >header();

echo "<h2 style=\"text align:center;\">Discussion Group Members</h2
    >";
echo '<table align="center" width="100%">';

foreach($gm as $gr ){
        echo '<tr style="border:1px solid gray;"><td style="border:1px
             solid gray;"><a href="individual.php?grd='.$gr >id.'&id='.
            $courseid.'" style="font weight:bold;">'.$gr >name.'</a></
            td>';
        $nouing = $DB >get_record_sql("SELECT count(u.username) as
            ucount FROM mdl_user u,mdl_groups_members gm WHERE gm.
            groupid = $gr >id AND u.id = gm.userid");
        $guser =  $DB >get_records_sql("SELECT u.username FROM
            mdl_user u,mdl_groups_members gm WHERE gm.groupid = $gr >id
             AND u.id = gm.userid");
  $i=0;

  $users= array();
    foreach($guser as $gruser){
            if ($i !=0) {
                        $users[]=$gruser >username;
                } else {
                    $users[]=$gruser >username. "<b>(Mentor)</b>";
```

```
            }
            $i++;

        }
        echo'<td style="border:1px solid gray;">';
        echo $userlist=implode(', ',$users);
        echo'</td>';
    echo '<td style="border:1px solid gray;">'.$OUTPUT >single_button
        (new moodle_url("/grouped/wekagroup.php" , array('grd'=>$gr >
        id , 'mail'=>'mail' ,'id'=>$courseid)), 'Mail')."</td>";
        echo '<td style="border:1px solid gray;">'.$OUTPUT >
            single_button(new moodle_url('/grouped/wekagroup.php' ,
            array('grd'=>$gr >id ,'sms'=>'sms', 'id'=>$courseid)), '
            sms')."</td>";
        echo "</tr>";

}
    echo "</table><hr />";
    echo "<h2 style=\"text align:center;\">iGCC: intelligent Grouping
        based on Collaboration Competence level</h2>";
    /* forms for message and sms*/
    if(isset($grd) && $grd!=0 ) {
        $results=$DB >get_records_sql("SELECT u.* FROM mdl_user u,
            mdl_groups_members gm WHERE gm.groupid = $grd AND u.id =
            gm.userid");
        $toaddress = array();
        $tonumbers =array();
        foreach ($results as $result){
                $toaddress[]  = $result >email;
                        $tonumbers[]=$result >phone1;
        }
        $names= $toaddress;
        $numbers=implode(',', $tonumbers);
}
```

```
if(isset($mail) && $mail =='mail') {
        echo $OUTPUT > box_start('generalbox');
        echo '<h3>Enter Subject and  Message to send email</h3></br
            >';
        echo '<form method="post"  action="wekagroup.php?id='.
            $courseid.'&grd='.$grd.'">
                                <table>
                                <tr><td><lable for="subject">Subject
                                    :</lable></td><td><input type="
                                    text" name="subject"></td></tr>
                                <tr><td><lable for="message">Message
                                    :</lable></td><td><textarea rows
                                    ="4" cols="50" name="message"></
                                    textarea></td></tr>
                                <tr><td></td><td><input type="submit"
                                     value="Submit" name="button1">&
                                    nbsp <button onclick="location
                                    .reload()">Cancel</button></td></
                                    tr>
                                </table>
                        </form>';
        echo $OUTPUT > box_end();
}

if (isset($_POST['button1'])) {
 $subject = $_POST['subject'];
 $message = $_POST['message'];
 foreach ($names as $to) {
  if(smtpmailer($to, $subject, $message)) {
   echo  ''. $to .'  Mail sent successfully! <br />';
  }
  else {
   echo'sorry mail not sent to' .$to. '</br>';
  }
 }
```

```
}

if(isset($sms) && $sms =='sms') {
  echo $OUTPUT >box_start('generalbox ');
  echo '<h3>Enter Message to send SMS</h3></br>';
  echo '<form method="post"  action="wekagroup.php?id='.$courseid.'&
      grd='.$grd.'">
  <table>
  <tr><td><lable for="message">Message:</lable></td><td><textarea
      rows="4" cols="50" name="message"></textarea></td></tr>
  <tr><td></td><td><input type="submit" value="Submit" name="button2
      ">  <button onclick="location.reload()">Cancel</button
      ></td></tr>
  </table>
  </form>';
  echo $OUTPUT >box_end();
}
if (isset($_POST['button2 '])) {
    $message = $_POST['message'].'</br>';
         message($message, $numbers);
}

  /**/

?>

<script type="text/javascript">
function validateForm()
{
var x=document.forms["myForm"]["noofclusters"].value;
  var y=document.forms["myForm"]["noofgroups"].value;
  var num_regex = /^\d+$/;
```

```
if (x==null || x=="")
  {
  alert("Number of Clusters must be filled out");
  return false;
  }
  if ( !x.match(num_regex) ) {
  alert("Number of clusters must be Integer");
  return false;
}
if (y==null || y=="")
  {
  alert("Number of groups must be filled out");
  return false;
  }

if ( !y.match(num_regex) ) {
  alert("Number of groups must be Integer");
  return false;
}
}
</script>
<!  Form to enter no of clusters to built     >
<form name="myForm" method="post" onsubmit="return validateForm()" >
    <h2>Create iGCC</h2>
  <table>
  <tr>  <td><label>Clustering Type</label></td>
                 <td> <select name="algtype">
                  <option>Skmeans</option>
                  <option>EM</option>
               </select></td></tr>
            <tr><td><label>Number of Clusters </label></td>
               <td><input type="text" maxlength="5"  name="
                   noofclusters" value="" /></td></tr>
                              <tr><td><label>Number of Groups</
                                 label></td>
```

```
                <td><input type="text" maxlength="5"  name="
                    noofgroups" value="" /></td></tr>
                        <tr> <td></td> <td><input type="submit" name
                            ="noofclrs" value="Next" /></td>
                            </tr>
                </table>
</form>

<! form ends >
<?php
                if(isset($_POST['noofclrs'])){
                $i = $_POST['noofclusters'];

//Executing java Class file
$file = 'data.csv';
if($_POST['algtype']=='EM'){
exec("java ClassesToClusters $i $file",$output);
} else {
exec("java MyKSVM $i $file",$output);
}

//Reading data from CSV file
                $row = 1;

//Deleting Existing Records from Temporary Folder
                $DB >delete_records("cluster_temp");
//Reading from output CSV  file which comes from weka
                if (($handle = fopen($output[0], "r")) !== FALSE) {
                        while (($data = fgetcsv($handle, 1000, ";"))
                            !== FALSE) {
                                    $num = count($data);
                                    $row++;

//Storing into data object
                                    $datas = new stdClass();
```

```
                    $datas >userID = $data[1];
                    $datas >clusterID = $data[2];

//Inserting data into cluster_temp table
                   $DB >insert_record('cluster_temp',
                      $datas);
              }// end of CSV data Read
                    fclose($handle);
        } //Closing CSV File

//noofusers , noofclusters
        $noofusers      = $DB >count_records_sql("SELECT
           count(userID) FROM {$CFG >prefix}cluster_temp");
        $noofclusters   = $DB >count_records_sql("SELECT
           count(distinct(clusterID)) FROM {$CFG >prefix}
           cluster_temp");
        $noofgroups = $_POST['noofgroups'];
//Displaying the results
              echo "<p>Total Number of Students: $noofusers
                 </p>";
              echo "<p>Total Number of clusters:
                 $noofclusters </p> ";
              for($i=0;$i<$noofclusters;$i++){
                    $cluster[] = $DB >count_records_sql("
                       SELECT count(clusterID) FROM {$CFG
                        >prefix}cluster_temp WHERE
                        clusterID=$i");
              }
              $clusterids = $DB >get_records_sql("SELECT
                 clusterID, userID from {$CFG >prefix}
                 cluster_temp ORDER BY clusterID ASC");
                foreach($clusterids as $clusterid){
                    $clus[] = $clusterid >clusterid;
              }
              $c=0;
```

```
                    foreach($cluster as $clusterno){
                            echo "<p>Number of Students in
                               Cluster$clus[$c] : $clusterno </p
                               >";
                            $c++;
                    }

//Displaying Cluster assignments table
echo "<h2>Clusters </h2>";
echo "<table >";
foreach($clus as $clusno){
echo "<tr><th style=\"border:1px solid gray;\">Cluster$clusno </th>";
$users = $DB >get_records_sql("SELECT u.username,u.id
FROM {$CFG >prefix}user u,{$CFG >prefix}cluster_temp ct WHERE
ct.clusterid = $clusno AND u.id = ct.userid");

foreach($users as $user){
echo "<td style=\"border:1px solid gray;\">$user >id</td>";
                            }
                            }
                            echo "</tr></table>";

//Storing cluster assignments in double dimentional array
                            $i=0;
                            foreach($clus as $clusno){

//$count =   $DB >count_records_sql   ("SELECT count(userID) FROM
{$CFG >prefix}cluster_temp WHERE clusterID=$clusno");
$userst= $DB >get_recordset_sql("SELECT u.firstname,u.id FROM
{$CFG >prefix}user u,{$CFG >prefix}cluster_temp ct WHERE
ct.clusterID = $clusno AND u.id = ct.userID");
                            $j=0;
                            foreach($userst as $user){
                            $arr[$i][$j] = $user >id;
                            $j++;
```

```
}
$i++;
}


//randomizing the  double dimentional array
if($_POST['algtype']=='EM'){
                    for($i=0;$i<$noofclusters;$i++){

                    $random_arr = $arr[$i];
                    shuffle($random_arr);
//randomize the array by shuffling it
                    $test[] = $random_arr;
                    };
} else{

                    for($i=($noofclusters 1);$i>=0;$i  ){

                    $random_arr = $arr[$i];
                    shuffle($random_arr);
//randomize the array by shuffling it
                    $test[] = $random_arr;
                    };
}
                     // $new = array();
                       for($a=0;$a<$noofgroups;$a++){
                       for($b=0;$b<sizeof($test[$a]);$b++){
                        $new[] = $test[$a][$b];
                        }
                        }
                       for($no=0;$no<$noofgroups;$no++){
                       for($rank=$no;$rank<sizeof($new);
                           $rank+=$noofgroups){
                                $group[$no][$rank] = $new[
                                    $rank];
```

```
                                }
                                }

//Display the groups table with UserIDs and GroupIDs
        echo "<h2>Groups</h2>";
        echo "<table cellpadding=\"5\">";
        //$clusn = 0;

       for($no=0;$no<$noofgroups;$no++){
        $f=0;
                $gnumber = $no+1;
         echo "<tr><th style=\"border:1px solid gray;\">Grouping(
            $gnumber)</th>";
        for($rank=$no;$rank<sizeof($new);$rank+=$noofgroups){
        echo "<td style=\"border:1px solid gray;\">";
        if($f!=0){
         echo $group[$no][$rank];
         } else {
           echo $group[$no][$rank]."(Mentor)";
          }
         echo "</td>";
         $f++;
         }
         echo "</tr>";
         //$clusn++;
                                }
echo "</table>";
//if($confirm ==1){
$gro = new stdClass();
$grmem = new stdClass();

//$gm = $DB>get_records_sql("SELECT id,name FROM mdl_groups WHERE
   courseid=$courseid");
//foreach($gm as $g){
//   $DB>delete_records_select('groups_members','timeadded!='.time()
```

```
    .'', array('groupid'=>$g >id));
//}
//
//$DB >delete_records_select('groups','timecreated!='.time().'',array
    ('courseid'=>$courseid));
$DB >delete_records('groups_temp');
$DB >delete_records('groups_members_temp');
for($no=0;$no<$noofgroups;$no++){
                    $gnumber = $no+1;
    $gro >name = "Group". $gnumber;
    $gro >courseid =$courseid;
$gro >timecreated = time();
$gro >timemodified = time();
$gids = $DB >insert_record('groups_temp',$gro);
for($rank=$no;$rank<sizeof($new);$rank+=$noofgroups){
    $grmem >groupid = $gids;
$grmem >userid = $group[$no][$rank];
$grmem >timeadded= time();
$gis = $DB >insert_record('groups_members_temp',$grmem);
}
}
// redirect($CFG >wwwroot."/grouped/wekagroup.php?id=$courseid");
}
echo "Want to contintue these
groups".html_writer::link(new moodle_url('/grouped/confirmgroup.php',
array('id'=>$courseid,'conf'=>1, 'sesskey'=>sesskey())), get_string('
    next'));

echo $OUTPUT >footer();

  if(!isset($_POST['noofclrs'])){
/* vars for export */
// database record to be exported
$db_record = 'mdl_summary';
```

```
// database variables
$hostname = $CFG >dbhost;
$user = $CFG >dbuser;
$password = $CFG >dbpass;
$database = $CFG >dbname;

// Database connecten voor alle services
$conn = mysql_connect($hostname, $user, $password)
or die('Could not connect: ' . mysql_error());

mysql_select_db($database)
or die ('Could not select database ' . mysql_error());

$query = "SELECT role.userid,count(if(post.parent=0 ,post.userid,NULL
    )) AS
Numberofpost,count(if(post.parent!=0 ,post.id,NULL)) AS
    Numberofreplies,
(select round(COALESCE(avg(rate.rating),0)) from mdl_rating AS rate
    where
post.id=itemid and rate.component='mod_forum' and rate.ratingarea='
    post') as avgrating
        FROM
        mdl_context AS context
        INNER JOIN mdl_role_assignments AS role
        ON role.contextid=context.id and role.roleid=5
        LEFT JOIN mdl_forum_posts AS post
        ON role.userid=post.userid
        WHERE
        context.instanceid=$courseid and context.contextlevel=50
        group by role.userid";
        $result = mysql_query( $query, $conn ) or die( mysql_error(
            $conn ) );
        $out ="";
    $file = fopen('data.csv','w');
```

```
        $field = mysql_num_fields($result);

// create line with field names
for($i = 0; $i < $field; $i++) {
   if($i==($field 1)){
         $out .=mysql_field_name($result,$i);
    } else {
  $out .=mysql_field_name($result,$i).',';
}
}
        $out .= "\n";
        // Add all values in the table
        while ($l = mysql_fetch_array($result)) {
  for($i = 0; $i < $field; $i++) {
     if($i==($field 1)){
        $out .=$l[mysql_field_name($result,$i)];
    } else {
  $out .=$l[mysql_field_name($result,$i)].',';
  }
  }

        $out .="\n";
        }

        // Output to browser with appropriate mime type
        fputs($file,$out);
        fclose($file);
        exit;
 }
  ?>

//RUN INFORMATION
=== Run information ===

Scheme:        weka.clusterers.SimpleKMeans  V  N 3  A "weka.core.
```

```
   EuclideanDistance  R first last"  I 500  S 10
Relation:     Book1_clustered weka.filters.unsupervised.attribute.
   Remove R1
Instances:    36
Attributes:   3
              post
              asses
              replies
Test mode:    evaluate on training data

=== Model and evaluation on training set ===


kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 2.598172668504364
Missing values globally replaced with mean/mode

Cluster centroids:
                            Cluster#
Attribute      Full Data          0          1          2
                    (36)        (9)        (6)       (21)
=========================================================
post                9.25     8.2222         16     7.7619
               +/ 4.6866  +/ 3.8006  +/ 4.3359  +/ 3.3898

asses             9.1944    10.3333    20.3333     5.5238
                +/ 7.877  +/ 6.7082  +/ 8.8015  +/ 4.3888

replies           1.3056     3.3333          2     0.2381
               +/ 1.4894  +/ 0.7071  +/ 1.0954  +/ 0.4364
```

Clustered Instances

```
0       9 ( 25%)
1       6 ( 17%)
2      21 ( 58%)
```

```
=== Run information ===

Scheme:       weka.clusterers.EM  I 100  N 3  M 1.0E 6  S 500
Relation:     Book1_clustered weka.filters.unsupervised.attribute.
   Remove R1
Instances:    36
Attributes:   3
              post
              asses
              replies
Test mode:    evaluate on training data

=== Model and evaluation on training set ===


EM
==

Number of clusters: 3


               Cluster
Attribute          0       1       2
               (0.29)   (0.5)  (0.22)
====================================
```

```
post
  mean          14.5007    7.053   7.3142
  std. dev.       3.923   2.9868   2.5167

asses
  mean          16.0397   4.3844  11.1594
  std. dev.      8.4961   3.2656   6.2708

replies
  mean           1.6802   0.2143    3.323
  std. dev.      1.2499   0.4104   0.7291

Clustered Instances

0       10 ( 28%)
1       18 ( 50%)
2        8 ( 22%)


Log likelihood: 7.43506

//
=== Run information ===SKmeans (151 students)

Scheme:       weka.clusterers.SimpleKMeans  V  N 3  A "weka.core.
    EuclideanDistance  R first  last"  I 500  S 10
Relation:     studyr1  weka.filters.unsupervised.attribute.Remove R1
Instances:    151
Attributes:   3
              post
              replies
              av_rating
Test mode:    evaluate on training data

=== Model and evaluation on training set ===
```

kMeans

======

Number of iterations: 6

Within cluster sum of squared errors: 8.026114852930109

Missing values globally replaced with mean/mode

Cluster centroids:

		Cluster#		
Attribute	Full Data	0	1	2
	(151)	(35)	(35)	(81)
post	6.9868	6.8571	12.3143	4.7407
	+/ 4.2646	+/ 2.809	+/ 3.428	+/ 2.867
replies	8.0132	9.9429	13.9143	4.6296
	+/ 7.5339	+/ 6.5212	+/ 9.8231	+/ 4.3458
av_rating	1.3709	3.6571	1.2857	0.4198
	+/ 1.5038	+/ 0.6835	+/ 0.9571	+/ 0.6683

Clustered Instances

0	35 (23%)
1	35 (23%)
2	81 (54%)

=== Run information ===EM (151 Students)

Scheme: weka.clusterers.EM I 100 N 3 M 1.0E 6 S 500

Relation: studyr1 weka.filters.unsupervised.attribute.Remove R1

```
Instances:     151
Attributes:    3
               post
               replies
               av_rating
Test mode:     evaluate on training data

=== Model and evaluation on training set ===


EM
==

Number of clusters: 3


              Cluster
Attribute           0        1        2
               (0.32)   (0.41)   (0.27)
=======================================
post
  mean        10.8014   5.4824   4.9878
  std. dev.    4.0936   3.4397   2.3785

replies
  mean        14.8949   4.0053    5.932
  std. dev.    8.6941   3.5055   4.3903

av_rating
  mean         2.0737        0   2.2912
  std. dev.    1.4254   1.5038   1.2475

Clustered Instances

0        44 ( 29%)
```

1 77 (51%)

2 30 (20%)

Log likelihood: 7.84619

=== Run information ===SKmeans(109 students)

Scheme: weka.clusterers.SimpleKMeans V N 3 A "weka.core. EuclideanDistance R first last" I 500 S 10

Relation: studyr2 weka.filters.unsupervised.attribute.Remove R1

Instances: 109

Attributes: 3

post

replies

av_rating

Test mode: evaluate on training data

=== Model and evaluation on training set ===

kMeans

======

Number of iterations: 7

Within cluster sum of squared errors: 4.084074225179686

Missing values globally replaced with mean/mode

Cluster centroids:

		Cluster#		
Attribute	Full Data	0	1	2
	(109)	(12)	(39)	(58)
post	5.8624	8.25	9.5897	2.8621
	+/ 4.213	+/ 4.4339	+/ 3.0584	+/ 1.9326

replies	9.9817	10.8333	16.6923	5.2931
	+/ 9.5403	+/ 5.5076	+/ 11.2511	+/ 5.3738
av_rating	0.3578	2.5	0.1795	0.0345
	+/ 0.8555	+/ 0.7977	+/ 0.4514	+/ 0.1841

Clustered Instances

0 12 (11%)

1 39 (36%)

2 58 (53%)

=== Run information ===EM(109 Students)

Scheme: weka.clusterers.EM I 100 N 3 M 1.0E 6 S 500

Relation: studyr2 weka.filters.unsupervised.attribute.Remove R1

Instances: 109

Attributes: 3

post

replies

av_rating

Test mode: evaluate on training data

=== Model and evaluation on training set ===

EM

==

Number of clusters: 3

	Cluster		
Attribute	0	1	2
	(0.57)	(0.14)	(0.29)

===================================

post			
mean	3.6284	8.7235	9.6198
std. dev.	2.5702	5.3614	2.258
replies			
mean	5.6396	14.7458	17.8608
std. dev.	5.0516	13.5328	7.9959
av_rating			
mean	0	2.0353	0.0776
std. dev.	0.8555	0.9271	0.2675

Clustered Instances

0	61 (56%)
1	14 (13%)
2	34 (31%)

Log likelihood: 7.06317