



**UNIVERSITY OF NAIROBI**  
**FACULTY OF SCIENCE AND TECHNOLOGY**  
**DEPARTMENT OF COMPUTING AND INFORMATICS**

**DEVELOPMENT AND EVALUATION OF A LEARNER BEHAVIOR MODEL FOR A LEARNING MANAGEMENT SYSTEM**

**BY**

**CHARLES LWANDE OMONDI**

**P80/98407/2015**

**SUPERVISORS: DR. LAWRENCE MUCHEMI, PROF. ROBERT OBOKO**

**MARCH 2022**

**Research submitted in fulfillment of the requirements for the award of the degree of  
Doctor of Philosophy in Information Systems**

# DECLARATIONS


I hereby declare that this thesis is my work and has, to the best of my knowledge, not been submitted to any other institution of higher learning.

Student: Charles L. Omondi (P80/98407/2015)


Signature:  Date: 7/3/2022

This thesis has been submitted as partial fulfillment of requirements for the award of the degree of Doctor of Philosophy in Information Systems of the University of Nairobi with my approval as the University supervisor.

Supervisor: DR. LAWRENCE MUCHEMI

Signature:  Date: 19<sup>th</sup> April 2022

Supervisor: PROF. ROBERT OBOKO

Signature:  Date: 19<sup>th</sup> Apr 2022

# **DEDICATION**

To my late parents who did not live to witness

## **ACKNOWLEDGEMENT**

I want to thank the University of Nairobi management for granting permission to carry out the research at the institution.

I wish to thank my supervisors Dr. Lawrence Muchemi and Prof. Robert Oboko for their guidance, thoughtful comments, and feedback during the research.

Special thanks to Dr. Paul Mungai of the department of medical physiology and Dr. Beda Olabu of human anatomy for their support during data collection with students. I thank the department of medical psychiatry for developing the rich learning content that provided the data for the research. I want to thank the preclinical students of the University of Nairobi for participating in this research.

I wish to thank my research assistants Margaret Nthenya Musembi, Purity Wambui Mwangi, Rufusy Idachi Ambasu, and my colleague Julian Akinyi Anyim.

Many thanks to the department of computing and informatics research team for providing many helpful comments and suggestions during seminar presentations.

Thanks also to my wife Ruth and the children Chrispine, Caro, Fabian, and Stanislaus for putting up with me being absent for hours on end.

Finally, I give glory to God for granting the energy and determination.

## **ABSTRACT**

Most Learning Management Systems (LMS) lack automated components that analyze data and generate information on learner behavior. Majorly traditional manual tools and methods such as the administration of questionnaires relating to a specific learning style and cognitive psychometric tests have been used to identify specific behavior. The problem with such methods is that a learner can give inaccurate information, is time-consuming, and is prone to errors. While there are existing automated models as reviewed in literature predicting learning style and cognitive traits, most of them are based on single behavior and tested on specific learning platforms.

The primary objective of this study was to design, develop and evaluate a model complementing Learning Styles with psychology-based ones such as Cognitive Traits. An automatic model based on Felder-Silverman Learning Style Model and Cognitive Trait Model was designed. Approximately 200,000 log records of 389 students who had accessed e-Learning courses for a 15-week semester were extracted from LMS to create a dataset. From this, a prototype estimating learning behavior based on the two theories was developed and evaluated with students in a classroom environment. The model can estimate learning behavior and display relevant learning content matching the learning styles and cognitive traits of an individual student. The evaluation of the model done using Kappa statistics demonstrated that the interrater reliability results were moderately in agreement with the traditional psychometric methods.

Further, machine learning methods involving artificial neural networks and cluster algorithms were used to analyze the dataset collected from the log records. The artificial neural network model was trained to predict learner behavior and classify each student. The prediction success rate of 0.63, 0.67, 0.64, 0.65, 0.26, and 0.64 for accuracies, precision, recall, f-score, kappa, and Area Under the Curve (AUC) respectively were recorded. On the other hand, the cluster evaluation analysis showed positive silhouette coefficients for all clusters indicating that all the students with similar browsing behavior had been assigned to the right clusters matching their learning styles and cognitive traits.

Cumulatively, these findings demonstrate that students with similar content navigation behavior share common learning styles and cognitive traits. These results suggest that it is possible to automatically estimate the learning styles and cognitive traits of a learner in a learning management system with fair accuracy. This research brings forth a generic modeling architecture that developers can integrate with existing learning management system platforms to improve learner characterization. Furthermore, a course lecturer can use the information generated by the model to provide learning materials matching identified characteristics for each student and also apply appropriate teaching methods.

**Keywords:** Learning style, Cognitive style, Learning management system, Learner modeling, Learner behavior, Machine learning, Neural network, Cluster analysis

## **LIST OF ABBREVIATIONS**

ALA – associative learning ability

ANN – artificial neural network

AUC – area under the curve

CT – cognitive traits

CTG – cognitive traits generator

CTM – cognitive traits model

DFD – data flow diagram

DIANA – Divisive Analysis

DL Distance Learning

EMIS - Education Management Information Systems

ERD – entity relationship diagram

FMMCasCorNN - Fuzzy Min-Max Cascade Correlation Neural Network

FSLSM - Felder-Silverman Learning Style Model

GPL – Gnu Public License

HBDI - Herrmann Brain Dominance instrument

ICT – information communication technology

ILS – index of learning style

ILMS - Intelligent Learning Management System

IPS – information processing speed

IRA – inductive reasoning ability

KNN – k-nearest neighbor

LMS – learning management system

LS – learning style

LS: CY - Learning Style: The Clue to You

LSG – learning style generator

LSI - Learning Style Inventory

LSQ - Learning Style questionnaire

MBTI - Myers Briggs Type Indicator

MCQ – multiple-choice questions

OLMS - Online Learning Management System

OSPAN - Operation Span

PAM - Partition around Medoids

PHP – hypertext preprocessor

SCORM - Sharable Content Object Reference Model

SQL – structured query language

VARCK - visual, aural, read/write, and kinesthetic

VAT – visual assessment tendency

VLE - Virtual Learning Environments

WEKA - Waikato environment for knowledge analysis

WMC – working memory capacity

# TABLE OF CONTENTS

DECLARATIONS .....	i
DEDICATION .....	ii
ACKNOWLEDGEMENT .....	iii
ABSTRACT .....	iv
LIST OF ABBREVIATIONS .....	v
LIST OF TABLES .....	xii
LIST OF FIGURES .....	xiii
LIST OF APPENDICES .....	xiv
CHAPTER 1: INTRODUCTION AND BACKGROUND INFORMATION .....	1
1.1 Learner Behavior .....	2
1.2 Learning Management Systems (LMS) .....	4
1.2.1 Features of a Learning Management System .....	4
1.2.2 The state of LMS in Kenya.....	6
1.2.3 Limitations of LMS .....	7
1.3 Problem Statement .....	8
1.4 Research Purpose .....	9
1.5 Research questions .....	10
1.6 Research Significance .....	10
1.7 Thesis outline .....	11
CHAPTER 2: LITERATURE REVIEW .....	12
2.1 Information Processing Theory .....	13
2.1.1 Models of Information Processing Theory .....	13
2.1.2 Information Processing Theory in the Context of Information Systems .....	16
2.1.3 Information Processing Theory in the Context of Computer-Based Learning.....	16
2.1.4 LMS in the Context of Information Processing Theory .....	18
2.1.5 The Weaknesses of an LMS in the Context of Information Processing Theory .....	20
2.2 Learning Style Models .....	22
2.2.1 Felder-Silverman Learning Style Model .....	23
2.2.2 Honey and Mumford Learning Style Model .....	24
2.2.3 Myers Briggs Type Indicator (MBTI) .....	24



2.2.4 VARK Learning Style Model.....	25
2.2.5 Kolb’s Experimental Model .....	26
2.2.6 Dunn and Dunn’s Learning Style Model.....	26
2.2.7 The Herrmann Brain Dominance Instrument (HBDI).....	27
2.3 Cognitive Traits Important for Learning.....	28
2.3.1 Working Memory Capacity .....	28
2.3.2 Inductive Reasoning Ability.....	29
2.3.3 Associative Learning Ability.....	30
2.3.4 Information Processing Speed.....	30
2.4 Learner Modeling.....	32
2.4.1 Gathering Data for Modeling .....	33
2.4.2 Learner Adaptation.....	34
2.5 Learner Behavior Modeling in an LMS .....	35
2.5.1 Learner behavior data that can be extracted from an LMS .....	35
2.5.2 Associating LMS Behavioral Patterns with LS and CT models .....	37
2.6 Estimating LS and CT from Behavior in an LMS .....	42
2.6.1 Machine Learning Methods.....	42
2.6.2 Estimation by Analyzing Browsing Patterns Using Rules .....	49
2.6.3 The gap in the Previous Studies .....	50
2.7 Theoretical Framework .....	53
2.8 Conceptual Framework .....	54
2.8.1 Related Study.....	54
2.8.2 Definition of Conceptual Framework in the Current Study .....	56
2.8.3 Operationalization of Variables.....	59
2.9 Summary .....	62
CHAPTER 3: RESEARCH METHODOLOGY .....	64
3.1 Research Design.....	64
3.1.1 The Population of the Study .....	66
3.1.2 Sample Selection .....	67
3.1.3 Data Collection.....	68
3.1.4 Model Development .....	68

3.1.5 Model Evaluation .....	69
3.1.6 Validity and Reliability of Model Evaluation Instruments .....	71
3.1.7 Presentation of the Evaluation Results .....	72
3.2 Learning Contents Hosted on LMS.....	74
3.2.1 Patterns Extraction from contents.....	75
3.2.2 Patterns Extracted from an LMS Log Records.....	78
3.3 Learner Behavior Model .....	79
3.3.1 LMS Log.....	81
3.3.2 Pattern extraction component .....	81
3.3.3 LS and CT Estimation Components .....	82
3.3.4 Content Matching Component.....	85
3.3.5 Display interface.....	87
3.4 Model Prototyping.....	87
3.4.1 Requirement Specifications .....	88
3.4.2 Prototype Development .....	89
3.5 Model Evaluation .....	95
3.5.1 Index of Learning Style Questionnaire.....	96
3.5.2 Online Cognitive Tests .....	96
3.5.3 Data Collection and Analysis Procedure .....	98
3.5.4 Interrater Reliability Analysis .....	100
3.6 Learner Modeling using Artificial Neural Network Analysis.....	102
3.6.1 Data Encoding .....	103
3.6.2 Data Pre-processing.....	109
3.6.3 Training Neural Network.....	110
3.6.4 Network Evaluation.....	115
3.7 Learner Behavior Modeling using Clustering Algorithms.....	116
3.7.1 Formulating Euclidean Distance .....	119
3.7.2 Cluster Analysis and Evaluation.....	122
3.7.3 Incorporating Clusters in the CogniLearn Model.....	124
3.8 Summary .....	125
CHAPTER 4: RESULTS AND DISCUSSIONS .....	127

4.1 Results for Prototype Evaluation.....	127
4.1.1 Descriptive Statistics .....	127
4.1.2 Evaluation Results for LS Prediction .....	131
4.1.3 Evaluation Results for CT Prediction.....	134
4.2 Evaluation Results for the Neural Network Model.....	138
4.2.1 Evaluation Results for all variables .....	139
4.2.2 Evaluation Results Variables Related to LS.....	141
4.2.3 Evaluation Results for Variables Related to CT.....	143
4.3 Evaluation Results for Cluster Models.....	145
4.3.1 Results for Variables Related to CT .....	146
4.3.2 Results for Variables Related to LS .....	147
4.3.3 Results for Hybrid Variables .....	149
4.4 Discussion of Results .....	151
4.4.1 Discussion of Evaluation Results for the Proposed Model .....	151
4.4.2 Discussion of Evaluation Results for the Neural Network Model .....	157
4.4.3 Discussion of Evaluation Results for the Cluster Analysis .....	159
4.5 Comparison of the results regarding the three approaches .....	161
4.6 Summary .....	163
CHAPTER 5: CONCLUSION AND RECOMMENDATIONS .....	165
5.1 Summary of the Findings .....	165
5.1.1 The available methods that are applicable for the automatic detection of learner behavior .....	165
5.1.2 How Feature Sets for Automatic Detection of Learner Behavior can be Extracted from LMS logs .....	167
5.1.3 How a Computing Model can be used to Identify Learner Behavior in an LMS.....	169
5.1.4 How the Performance of the Model can be Evaluated .....	171
5.2 Research Contributions .....	173
5.2.1 Practical Contribution.....	173
5.2.2 Methodological Contribution .....	173
5.2.3 Theoretical Contribution.....	175
5.2.4 Contribution to the Field of Information Systems.....	176

5.3 Limitations of the Current Study.....	177
5.4 Recommendations for Further Research Work.....	177
REFERENCES .....	178
APPENDICES .....	193

## LIST OF TABLES

Table 2.1. Operationalizing Variables related to FSLSM.....	59
Table 2.2. Operationalizing Variables related to CT .....	61
Table 3.1. Sample Log Data Extracted from an LMS .....	79
Table 3.2. LS Estimation functions.....	83
Table 3.3. CT Estimation Functions .....	85
Table 3.4. Matching the Learning Preferences with Appropriate Contents .....	86
Table 3.5. Interrater Reliability Test.....	100
Table 3.6. Sample Dataset for Log data extracted from an LMS .....	104
Table 3.7. Input and output variables.....	106
Table 3.8. Sample Dataset .....	107
Table 3.9. Frequency Distribution for Learner Behavior .....	108
Table 3.10. Computing Euclidean Distance Between a Pair of Students .....	119
Table 3.11. Allocating Content to Students with Unknown LS and CT.....	124
Table 4.1. Distribution of Learning Styles by Strengths .....	128
Table 4.2. Distribution of ALA by Strengths .....	129
Table 4.3. Distribution of IRA by Strengths.....	130
Table 4.4. Distribution of WMC by Strengths.....	130
Table 4.5. Distribution of IPS by Strengths.....	131
Table 4.6. Interrater Reliability Analysis Results for Active-Reflective LS .....	131
Table 4.7. Interrater Reliability Analysis Results for Sensing – Intuitive LS .....	132
Table 4.8. Interrater Reliability Analysis Results for Sequential – Global LS.....	133
Table 4.9. Interrater Reliability Analysis Results for Associative Learning Ability.....	134
Table 4.10. Interrater Reliability Analysis Results for Information Processing Speed.....	136
Table 4.11. Interrater Reliability Analysis Results for Inductive Reasoning Ability .....	137
Table 4.12. Interrater Reliability Analysis Results for Working Memory Capacity .....	138
Table 4.13: Evaluation Results for all LS Variables.....	141
Table 4.14. Evaluation Results for LS Variables Related to Content Pages Visited.....	142
Table 4.15. Evaluation Results for LS Variables Related to Time Spent on the Content Pages	142
Table 4.16: Evaluation Results for Variables Related to Cognitive traits .....	143
Table 4.17: Summary of Evaluation Results .....	144
Table 4.18. Silhouette Analysis Results for Variables Related to Cognitive Traits.....	146
Table 4.19: Cluster analysis results for variables related to LS.....	147
Table 4.20: Cluster Analysis Results for Variables related to LS and CT .....	149
Table 4.21. Evaluation Results for the Cognilearn Model.....	162
Table 4.22. Evaluation Results for the ANN Model.....	162
Table 4.23. Evaluation Results for the Cluster Model.....	162

## LIST OF FIGURES

Figure 2.1 Conceptual Framework .....	58
Figure 3.1. Research Process .....	66
Figure 3.2. CogniLearn Model Architecture.....	80
Figure 3.3. Matching Content with Relevant LS and CT .....	85
Figure 3.4.1 LS Estimation Algorithm .....	91
Figure 3.4.2. CT Estimation Algorithm.....	93
Figure 3.5. Neural Network Architecture .....	102
Figure 3.6. Example of LS and CT Output Generated .....	106
Figure 3.7. Response variable distribution. ....	110
Figure 3.8.1. Neural Network Model Indicating all Input Variables.....	113
Figure 3.8.2. Neural Network Model Indicating LS Input Variables Related Content .....	114
Figure 3.8.3. Neural Network Model Indicating LS Input Variables Related to Time .....	114
Figure 3.8.4. Neural Network Model Indicating Input Variables Related to CT .....	115
Figure 3.9. The Euclidean distance the matrix generated. ....	120
Figure 3.10. The visual assessment of the distance matrix.....	121
Figure 4.1. Distribution of Participants by Gender (N=124).....	128
Figure 4.2.1. Confusion matrix table. ....	140
Figure 4.2.2. Area Under the Curve (AUC).....	140
Figure 4.3. Silhouette Plot for Variables Related to CT ( k=10) .....	147
Figure 4.4. Silhouette Plot for Variables Related to LS ( k=10) .....	149
Figure 4.5. Silhouette Plot for Hybrid Variables (k=10) .....	150

## LIST OF APPENDICES

Appendix 1. Publications .....	193
Appendix 2. Letter of Request to Conduct Research at the University .....	194
Appendix 3. The Electronic Learning Module Sections hosted on Claroline LMS .....	195
Appendix 4. A Sample Online SCORM Content on Claroline LMS .....	195
Appendix 5. The Content Navigation Statistics generated from Claroline LMS logs.....	196
Appendix 6. Sample MCQ hosted on Claroline LMS (source: The University of Nairobi LMS) .....	196
Appendix 7. Sample Mark Sheet generated from Claroline LMS.....	197
Appendix 8. Sample Raw Data from LMS Logs.....	197
Appendix 9. The System Flow Diagram for Cognilearn Model.....	199
Appendix 10. Evolutionary Prototyping Process.....	200
Appendix 11. Model Entity-Relationship Diagram (ERD for the Cognilearn Model Database. 200	
Appendix 12. Data Dictionary for the CogniLearn Model Database. ....	201
Appendix 13. CogniLearn Model Program Code Segments .....	201
Appendix 14. Model Interface and Functionalities.....	208
Appendix 15. Index of Learning Styles (ILS) Questionnaire and the scoresheet.....	214
Appendix 16. Survey data collected from ILS questionnaire.....	216
Appendix 17. Classifications of survey results collected from ILS questionnaire.....	219
Appendix 18. Cognitive Assessment Tests.....	222
Appendix 19. Classification of survey results for psychometric tests related to ALA.....	224
Appendix 20. Classification of survey results for psychometric tests related to IRA.....	226
Appendix 21. Classification of survey results for psychometric tests related to IPS.....	229
Appendix 22. Classification of survey results for psychometric tests related to WMC.....	232
Appendix 23. Classification of results predicted for active reflective LS.....	235
Appendix 24. classification of results predicted for sensing intuitive LS.....	238
Appendix 25. classification of results predicted for the sequential global LS.....	241
Appendix 26. Classification of results predicted for the visual verbal LS.....	244
Appendix 27. Classification of results predicted for the ALA.....	245
Appendix 28. Classification of results predicted for IRA.....	250
Appendix 29. Classification of results predicted for the IPS.....	253
Appendix 30. Classification of results predicted for the WMC.....	256
Appendix 31. Sample Training Dataset for the ANN and the cluster models.....	260
Appendix 32. Neural Network Program Code Segment.....	260
Appendix 33. Sample Prediction Results for the ANN model.....	264
Appendix 34. Cluster Analysis Program Code Segment.....	264
Appendix 35. Sample Cluster model results for 20 Students.....	271
Appendix 36. Pattern extraction queries.....	271

# **CHAPTER 1: INTRODUCTION AND BACKGROUND INFORMATION**

The recent past has witnessed a rapid development of information and communication technologies. These developments have placed demands on institutions of higher learning to innovate and operationalize management information systems that effectively address the needs of their customers, improving the institutional decision-making processes and efficient management to fully meet their goals. Management information systems provide strategies for an organization to reorganize the information used for decision-making to achieve its goals (Kroenke and Boyle, 2019; Mukhtar, Fransiska, and Wahyudi, 2018). When we talk of management information systems, it refers to software, hardware, people, and procedures (Williams and Sawyer, 2014). The information systems use hardware, software, and telecommunications network to collect, create, and distribute useful data that meet the needs of an organization (Gray et al., 2016; Valacich and Schneider, 2015). Today many organizations are even applying knowledge management in managing information systems (Chulkov, 2017; Adebayo and Wokocha, 2011). Knowledge management is the development of methods, tools, and techniques that facilitate the flow of information between users and its retrieval, processing, and use of such information in improving and innovating organizational activities (Gonzalez and Martins, 2017).

In the context of education, there are Education Management Information Systems (EMIS) which are used to collect, integrate, process, maintain, and disseminate data and information to support decision-making, planning, analysis, monitoring, and evaluation of all education system levels (Borkar, 2022; Helal, Ahmed, and Bhuiyan, 2021). The use of EMIS in institutions of learning as has been evidenced in literature sources has highlighted many advantages including better information access, efficiency in administration, better attendance records, efficient examination



management, improved learning services, and improved quality of reports (Ugwude and Ugwude, 2020; Kuswara *et al.*, 2019; Shah, 2013). EMIS gives an institution of learning overall reliable and timely information for monitoring and evaluating strategies (Bojte, 2019). However, any effective EMIS must bear in mind that students have different cognitive behavior that determines the way they understand the content.

The internet has made possible dissemination of knowledge through Virtual Learning Environments (VLE) and enhanced access to distance learning (DL) programs which has seen their adoption in institutions of learning around the world (Allen and Seaman, 2017). VLEs are online EMIS platforms that allow synchronous and asynchronous communication in a learning environment contributing to the generation of learning resources that can be used to teach students (Andrade, Rigo, and Barbosa, 2021; Waheed *et al.*, 2020). VLEs are often referred to as Learning Management Systems (LMS). This thesis will use the term Learning Management System.

As a result of the above observations, studies, investigating how students learn and acquire knowledge in an online environment have attracted the attention of researchers in the recent past. This chapter provides the background information for the study. It outlines the concept of learner behavior and learning management systems (LMS). It also states the problem statement, the research questions, and the significance of the study. It ends by providing the overall structure of the thesis.

## **1.1 Learner Behavior**

Research has shown that students have significant categories of diversity in learning behaviors that influence their successful learning. Some of the diversities are student learning styles, levels of intellectual development, and how they approach learning (Felder, 1988). According to Felder, students learn by visualizing, listening, participating, logical reasoning, memorizing, and

remembering. Further, Felder emphasized that students experience learning challenges when the environment does not support their individual needs.

According to Eliezer *et al.* (2022), learning achievements improve when one is aware of own learning preferences and uses them efficiently. The study indicated that every learner has more than one learning preference and they vary for both males and females. It further alluded that, a person who knows their weaknesses in learning better decides which learning preference is suitable to apply. This increases achievements, self-confidence, and learning attitude. Therefore identifying learner behavior is an important process in understanding and improving the learning experience. This is achieved through learner behavior modeling.

Learner behavior modeling is the process of collecting and processing data about learner preferences during a learning process (Brusilovsky and Millán, 2005).

Understanding student learning styles may have a dramatic influence on the success of content delivery (Goodridge, Lawanto, and Santoso, 2017). Hamed and Almabruk (2021) reported a strong correlation between students' learning preferences and their academic achievements. Hamed and Almabruk recommended that instructors should identify perceptual learning behavior and match them with their teaching strategies to influence students' achievements.

Andersen and Sorensen (2017) emphasize that knowing learning preferences guide an instructor to assign learning tasks suitable for individual learning abilities. Furthermore, a content designer can prepare learning materials suitable for learners' cognitive abilities and learning preferences (Chang *et al.*, 2016; Yen *et al.*, 2015)

## **1.2 Learning Management Systems (LMS)**

Learning Management System (LMS) is a web-based platform that facilitates the delivery of instructional materials, assignments, and collaboration between students and instructors (Ulker and Yilmaz, 2016). An LMS is implemented and hosted on a central server to facilitate teaching online (Baumgartner, Häfele, and Maier-Häfele, 2002). Many other terms exist in the literature with similar meanings e.g. Course Management System, Learning Management System e-Learning platform, and Virtual Learning Environment.

LMS is used by institutions of learning such as schools, colleges, universities, and training providers to manage online courses. An LMS can be deployed on-premises and in the cloud (Bouchrika, 2022a).

Some of the most well-known open-source LMS platforms are Claroline (*Claroline*, 2017), Moodle (*Moodle*, 2018), Ilias (*Ilias*, 1998), and ATutor (*ATutor*, 2002). These platforms are freely distributed under appropriate creative commons licenses (*Creative Commons*, 2013).

In Kenya, most universities are using Moodle LMS. The University of Nairobi until recently, in the year 2020, was using the Claroline platform. It currently utilizes both Moodle and Claroline LMS platforms which are implemented on Hypertext Preprocessor (PHP) scripting language and uses Apache and MySQL as web and database servers respectively.

### **1.2.1 Features of a Learning Management System**

The key aspect of an LMS is the ability to provide virtual classroom features for the learners. These features enable an instructor to manage content, assignments, and online tests. It as well facilitates the management of users and courses. Course contents can be arranged in a different order to suit the preference of an instruction designer. Each course can be divided into lessons or sections in

which learning resources e.g. handouts, video files, external links, and notes can be created. Learning activities like quizzes, forums, exercises, and assignments can also be created within a course in an LMS.

Some of the basic common features in most LMS include:

- Course management – This feature facilitates content creation and delivery. The course management tools allow an instructor to upload content, tests, illustrations, and videos and create links to external resources. In addition, an instructor can arrange learning materials to create a sequence of learning activities. The content management feature has calendar tools for creating and publishing scheduled events.
- User management - this feature is used to organize students. The students can be organized into groups based on specific learning criteria. This enables an instructor to assign learning activities tailored to each group.
- Reporting and analytics – The systems generate reports on learners' online activities and course participation. These include records on course access frequencies, time spent, and assessments. This feature provides reports indicating the online navigation behavior of a student. The feature also provides data on course learning progress, the frequency of content visits, and time spent. Some analytic reports present information in graphical form.
- Assessment - this feature allows an instructor to create assessment tests. Most LMS allows the creation and automatic grading of multiple-choice tests. An instructor can use the feature to create tests at the end of each chapter to make learning easy.
- Discussion – Most LMS have collaborative tools that enable learners to work together using forums, wikis, and glossaries. These features allow students to collaborate by sharing

knowledge. An instructor can utilize this feature to facilitate online engagement and consultations with students.

### **1.2.2 The state of LMS in Kenya**

The increased demand for university education in recent years has led to significant growth in online learning in Kenyan universities. Most institutions of higher learning in Kenya have embraced online learning to cope with the ever-increasing student population. This has seen the adoption of various LMS platforms by different institutions of learning to facilitate the remote learning process. This phenomenal development of online learning has come with its share of implementation challenges in efforts to fully utilize LMS to support teaching and learning.

In their investigation to establish the status of e-learning in public universities in Kenya, Makokha and Mutisya (2016) reported that a few lecturers and students used LMS. The study further pointed out that a few courses were on LMS with just non-interactive lecture notes uploaded to the system. In this study, Makokha and Mustiya concluded that insufficient time for online interaction was the most serious challenge affecting the use of LMS in these institutions of higher learning.

These sentiments were recently supported by Kibuku, Ochieng, and Wausi (2020) who argued that the lack of technical and pedagogical competencies and training for e-tutors and e-learners, lack of an e-Learning theory to underpin the e-Learning practice were the challenges faced in the implementation and provision of e-Learning in universities in Kenya. In a previous study, Tarus and Gichoya (2015) suggested that technological, organizational, and pedagogical components are the preconditions necessary for the successful implementation of e-learning in Kenyan universities.

### **1.2.3 Limitations of LMS**

Even though Learning Management Systems (LMS) have been used by learning institutions to deliver content, these platforms lack an intelligent mechanism for analyzing browsing records and classifying learners according to their respective learning behavior. Most of the time, an LMS only produces daily access reports to show the number of hits recorded on various parts of an LMS during various time frames. These reports are only useful for administrative purposes but do not help instructors to draw useful learner behavior.

Subsequently, there has been ongoing research in the context of creating an Intelligent Learning Management System (ILMS) that uses artificial intelligence to extract information hidden in LMS logs through learning analytics (Romero and Ventura, 2020). For example, some authors have been interested in analyzing records of user interactions with an LMS to predict performance (Chango *et al.*, 2021; Chen and Cui, 2020; Naeim *et al.*, 2022).

Other studies have focused on ILMS that apply artificial intelligence methods to provide tutoring functions on an LMS e.g. mathematics (Vaerenbergh and Perez-Suay, 2022; Vaerenbergh and Perez-Suay, 2018). Some studies collect and analyze data on student interactions with an LMS to improve tutoring actions (Kurvinen *et al.*, 2020). These studies tend to focus on applying educational artificial intelligence techniques e.g. machine learning to conduct scheduled assessments of student activities in an LMS while ignoring personality traits that influence a learning process. What is not clear from the studies is the ability of these models to classify learners in terms of their respective psychological learning behavior. This indicates the need for a further study to understand learner preferences during online learning.

### **1.3 Problem Statement**

Most automated models for identifying learner behavior in an LMS only investigate a single psychological characteristic e.g. identifying learning styles (Rajper *et al.* 2016; Chang et al. 2016) or cognitive traits (Poellhuber and Roy, 2019; Lee and Natriello, 2018; Kamsa, Elouahbi, and El Khoukhi, 2018) but not both. A major weakness with such a learner behavior model is an incomplete or partial learner modeling process and thus unreliable information that cannot guide an instructor to apply appropriate teaching methods. These models, despite the good efforts, still have not exhaustively employed methods that investigate multiple psychological learning behavior in an LMS.

There are a few automated models for identifying multiple learner behavior in an LMS e.g. classifying students according to their anxiety, personality, and learning preferences (Aditya and Babita, 2018), identifying learning styles and affective states (Khan *et al.*, 2019). However, the techniques applied to identify feature sets from LMS logs and relate them to different psychological theories that describe learning behavior is still inconsistent. There is little agreement to date on how to infer learner behavior from the records of interactions with an LMS e.g. the learning object visited by a student (Correa, Monsalve, and Tabares, 2020), and participation in the discussion forum (Renato, 2020). There is still a lack of universal terms for relating learner behavior patterns in an LMS to psychological theories.

In addition, the automated models cited in the literature are developed and tested on specific learning platforms such e.g. Moodle (Villegas-Ch et al., 2018) thus results cannot be generalized to apply to other environments.

These automated models apply different classification methods to identify the learning behavior e.g. identifying learning styles using a decision tree algorithm (Liyanage, Gunawardena, and Hirakawa 2016), estimating learning styles using agents (Chang et al. 2016) and Bayesian networks (Rajper *et al.* 2016).

Due to these inconsistencies, further in-depth research is needed to allow a more accurate understanding of the relationships between learning behavior in an LMS and psychological theories. (Assis *et al.*, 2022)

The focus of this thesis is to suggest an automated method for inferring students' learning styles and cognitive traits from an LMS. This brings forth an automated hybrid model that will lead to better learner characterization in an LMS.

#### **1.4 Research Purpose**

This study sought to remedy the above problem by proposing an automated hybrid learner behavior model by applying the learning style and cognitive traits models. The two models form critical learner modeling theories which encapsulate personality characteristics. Personality traits play the most important roles in processing and organizing information. Learning style defines how a student acquires knowledge and skills. Cognitive trait defines how an individual perceives information. For example, some students understand better when they visualize content while others like listening to audio materials. Other students learn well when they discuss while others work alone. This research study adopts a multidisciplinary approach, comprising information systems, psychology, and instruction design which are essential for its successful outcome.



## **1.5 Research questions**

This research addressed the following questions:

- 1** What are the available methods applicable for the automatic detection of learner behavior in an LMS?
- 2** How can the feature sets for the automatic detection of learner behavior be extracted from LMS logs?
- 3** How can a computing model be used to identify a hybrid learner behavior in an LMS?
- 4** How can the above computing model be evaluated?

## **1.6 Research Significance**

This study contributes to the growing area of research in adaptive education systems by presenting a hybrid learner model which gathers data about students' characteristics and actions from an LMS to calculate students' learner behavior. The current study uses a rule-based system and machine learning methods to estimate multiple traits. The information generated by the model provides a better holistic representation of learner behavior that an instructor can rely on to apply appropriate teaching methods. These findings would make important contributions in the field of learning analytics on how to extract information hidden in LMS logs. The study presents a method of identifying feature sets from LMS logs that relate to different theories that describe learner behavior.

## **1.7 Thesis outline**

The thesis is presented in five chapters. Chapter one presents background information.

Chapter two addresses available techniques for identifying learner behavior in an LMS. It also provides a discussion on information processing theory and the related models. The first section of the chapter discusses personality models applied to LMS. The next section takes a look at the learner modeling process and the modeling data. The third section examines the modeling techniques in use. Finally, the last section presents the conceptual framework.

Chapter three discusses the modeling methods used in this study. The first section describes the research design. The second section discusses the feature extraction method. The third section outlines the modeling methods used in the study to estimate the learning preferences from the extracted features.

Chapter four presents the findings which emerged from the modeling process. The first section discusses the evaluation results of the proposed learner behavior model. The second section provides discourse on evaluation results for the neural network model. It ends by providing discussions on the evaluation results for cluster analysis.

Chapter five discusses the implication of the findings to future research in this area. The first section presents the summary of the findings. The second part discusses the significance of the findings from this study. The next part outlines the limitations of this study. The study concludes by making recommendations on areas for further research.

## **CHAPTER 2: LITERATURE REVIEW**

Learner modeling in a Learning Management System (LMS) is an integral part of the system. It has attracted increasing educational psychologists' attention forming an important research area to understand learner behavior with the main objective to foster meaningful teaching and learning in an online environment. While appreciating the existing large volume of published studies in this area, the majority have mainly focused on single-behavior modeling. The key question in this chapter is: what are the available methods for the automatic detection of learner behavior in an LMS? To answer this question, this study reviewed relevant published work in this area of study. Journal articles and conference papers as secondary sources of data were mined to identify the available methods and any research gaps present.

The section begins by presenting the available psychological theories and their implications on pedagogy. The information processing theory and its implications in information systems are discussed. The weaknesses of an LMS in the context of information processing theory are pointed out.

Issues regarding the incorporation of information processing theory in an LMS are discussed. The specific learning styles and cognitive traits models are discussed. Then the examples of specific indicators of learning styles and cognitive traits identified in the previous are presented. The techniques for estimating learner behavior applied in the related studies are highlighted.

Finally, the gaps in the previous studies are highlighted and a hybrid modeling method incorporating learning styles and cognitive traits is proposed. The section concludes by presenting a conceptual framework highlighting the independent and dependent variables to illustrate the

relationship between the patterns of learner behavior in an LMS and the psychological theories investigated.

## **2.1 Information Processing Theory**

Studies in cognitive psychology during the past decades have provided explanations of how the mind thinks, reasons, and learns (Slate and Charlesworth, 1989). This has been attributed to the development of the information-processing model of human memory. The psychology model is a theoretical framework that explains the thinking, reasoning, and learning of human memory (Atkinson and Shiffrin, 1968; Craik and Lockhart, 1972; Anderson, 1985; Gagne, 1985). The theoretical framework views the functions of human cognition as similar to computer operations.

Slate and Charlesworth (1989) identified the main theoretical concepts related to attention as active learning, meaningfulness, organization, advance organizers, memory aids, overlearning, and automaticity. Biehler and Snowman (1986) emphasized that instructors needed to apply techniques for attracting and holding the attention of students to make them actively engaged in a learning process. The psychologists emphasized that learning contents were only meaningful if used and retained for longer periods instead of rote learning (Biehler and Snowman, 1986; Gagne, 1985; Brown, 1975). Instructors need to know that students process information and learn in different ways (Glover, Bruning, and Filbeck, 1983).

### **2.1.1 Models of Information Processing Theory**

Information processing theory offers useful explanations of how the mind perceives and remembers information (Slate and Charlesworth, 1989).

The stage theory is one of the models widely used to describe information processing (Atkinson and Shiffrin, 1968). The stage model specifies that information goes through sensory memory,

short-term memory, and long-term memory to become encoded in the mind. Sensory memory receives and transfers new information to the next processing stage. Short-term memory stores information for 15-30 seconds then loses other actions that are not taken. Long-term memory stores previous perceptions, knowledge, and information learned.

Information kept in the long-term memory can reside in a dormant state until fetched back into consciousness (Abbot, 2002). Long-term memory communicates with short-term memory and it is dynamic. Abbot further states that a declarative memory can be verbalized i.e. the amount of deposited information that can be recovered and converted into words in functional thought and sharing. Ability to remember the real situation or an incident is controlled by episodic memory. The ability to deal with general, abstract information that can be remembered independently of how it was learned is determined by semantic memory. Learned skills and the capacity to remember instruction-like memory describe procedural memory. The structure for gathering and keeping information related to visual elements defines an imagery memory. This type of memory records information like a photograph and is useful for visual presentation. These factors help psychologists explain how the mind operates.

There are also the Parallel-Distributed Processing Model and Connectionist Model (Rumelhart and McClelland, 1986) which argue that numerous components of the memory system process information simultaneously. This theory proposes that the simultaneous occurrence of distributed signal activities via connections can be represented numerically. The connection strengths are modified based on experience for learning to occur (Smolensky, 1999). An artificial neural network consisting of simple processing units and a set of connections can be used to represent human thinking. Signals between nodes are transmitted based on the connections, the strength of the signal depends on the activation values of the input nodes and the weights of the connections.

Connectionist models thus emphasize a common technique for processing words, decentralized representation, and weighted connections (Seidenberg and McClelland, 1989). Various storage locations in the brain are connected through a network. Information is stored in these locations. It is easier for an individual to retrieve information with more connections.

Craik and Lockhart's Level of Processing Model (Craik and Lockhart, 1972) indicated that the extent to which information is elaborated upon affects its accessibility in memory. Memory recall stimuli is a function of mental processing depth. Thus, the more elaborate, longer-lasting, and stronger memory traces are synonymous with deeper than shallow levels of analysis.

Dual Coding Theory (Paivio and Desrochers, 1981) suggested two separate processing systems which deal with verbal and non-verbal information. The theory states that different systems process mental images and verbal objects.

From the foregoing discussion, the information processing model informs the foundation of mental psychology which is applied in computer technology to illustrate the processing mechanism of the mind (Vinney, 2020). Understanding these theories enable educationists to deal with the emerging challenges brought by technology in this technological era. The undertaking of the study in information systems theory is that human memory operates in the same way as a digital computer. The psychological states and processes are similar to computational states and processes.

Learning styles models and cognitive trait models are two behavior patterns developed from information-processing theories. Cognitive traits models define how an individual organizes and represents information in the mind while learning style defines how a student prefers to learn (Riding and Rayner, 1998). Education psychologists propose that individuals learn and perceive information differently.

### **2.1.2 Information Processing Theory in the Context of Information Systems**

The Information Processing Model of Learning associates the human mind with the way that a computer takes in data, processes, and stores information (Gagne, 1985). The learning process is similar to data processing steps in information systems. The steps are data input, processing, and storage for later retrieval. The input stage captures data into a computer. The processing stage converts data to a meaningful form to produce information. The storage keeps the information for future retrieval. This is comparable to the way a human mind processes data from learning in sensory, short-term, working memory, and long-term (Betz, 2006).

According to Woolfolk (2004), the mind frequently gets data into the sensory memory. The mind then holds data for a few seconds in short-term memory. Short-term memory briefly keeps data before the working memory can accommodate the new learning into the long-term memory for permanent or long-term storage. Data in long-term memory can be retrieved after initial learning.

According to Bouchrika (2022), the information processing theory relates the mind to a computer because combining new with stored information reveals new information that can provide solutions to various problems. It takes note that the power of a computer's central processing unit just like human attention has a limited capacity.

### **2.1.3 Information Processing Theory in the Context of Computer-Based Learning**

According to Salomon and Perkins (1998), the structure of observed learning outcomes (SOLO) and social aspects characterize learning in an educational environment. These aspects must be addressed by information processing theory in online learning. Biggs and Collis (1982), indicated that the model structure of observed learning outcomes (SOLO) is a taxonomy that defines stages of growing difficulty in a learner's understanding.

Information processing theory assumes that the human brain and computer memory process information in a similar way. The working memory shifts focus from desirable education materials when presented with extraneous information due to cognitive overload. The cognitive load on their mind is closely associated with the information processing theory (Paas and Ayres, 2014). Thus an instructor writing online learning resources must be guided by information processing theory (Ruiji, 2012).

According to Kalyuga (2012), while digital resources presented in an LMS can enrich interactivity, they can also cause cognitive overload and make it impossible for the learner's memory to process information. These online instructional designers must reduce the amount of noise in the content to make efficient use of computer-based learning strategies.

Taylor (2013), in his study, emphasized that where visual aids and text are used in the same online instructional materials, the two should be integrated so that a learner does not have to refer to the text to get information from the image. This helps to avoid unwanted noise and cognitive overload. Combining text, visuals, and audio in the same instructional materials increases cognitive load (Clark and Mayer, 2011).

In a study conducted by Zeglen and Rosendale (2018) to explore the effects of embedded games in online learning on learners' information retrieval and learning, the researcher used an LMS and created a quest-style educational game. According to Zeglen and Rosendale (2018), students who were exposed to a visual hint and elaborate feedback scored significantly higher than those in the control group who had no access. The study concluded that the transmission of information from working memory to long-term storage was affected by elaborate feedback and visual hints. This confirms that giving students pedagogical reinforcement makes them more likely to retain



information and achieve the learning objectives. Visual reinforcements and delayed feedback in an LMS should be considered to improve learning outcomes in an online environment. In addition, instructional designers should take advantage of technological features using reliable pedagogical approaches to enhance online learning. The use of delayed elaborate feedback and visual hints in an online environment has significant effects on a student's ability to process information from working memory capacity to long to memory. Thus organizing and presenting learning resources in a familiar, engaging, and motivating environment can provide can accelerate the transfer of information from a learner's working memory to long-term storage.

Zeglen and Rosendale (2018) however, suggested that practitioners should work together to further investigate the use of specific learning strategies in serious educational environments.

But still, many of the studies cited in this section have focused on the advantages of visual games in online learning and not on the didactic effect the game design has on the desired psychological learning preferences for each specific user in a specific course.

This thesis furthers the studies by investigating the type of information needed from students to identify their learning behavior in an LMS. The study presents hybrid models for identifying learner behavior based on multiple psychology theories that define learning characteristics. The proposed model filters and presents learning materials relevant to individual psychological characteristics to avoid subjecting learners to cognitive overload.

#### **2.1.4 LMS in the Context of Information Processing Theory**

There are specific areas where an LMS can be used to facilitate teaching and learning irrespective of the platform implemented by an institution (Baumgartner, Häfele, and Maier-Häfele, 2002).

LMS can also facilitate interactive communications through forums, chat, and video conferencing.

LMS also allows teachers to carry out evaluations and assessments of student performance using the assignments and exercises features. The system as well supports issues related to learner administration such as monitoring learning progress. Generally, the features available on LMS are meant to support cognitive strategies defined in learning theories.

Most LMSs are built based on pedagogical strategies that guide how learning takes place on the platform. Pedagogy explains how knowledge and skills are communicated based on an interaction between an instructor and a learner during a learning process (Watkins and Mortimore, 1999). These strategies are based on learning theories such as constructivism, constructionism, social constructivism, connected and separated behavior.

- Constructivism argues that learning is an active process where learners construct new knowledge according to their current and past knowledge during interaction with the environment.
- Constructionism argues that learning makes sense when a student is constructing knowledge for fellow students to learn.
- Social constructivism promotes constructivism by promoting social aspects that arise when learners work in groups.
- Connected and separated behavior promotes the motivation of students in a discussion. This promotes holding objective and factual arguments while listening to other learners' points.

For example, the design and development of open-source LMS Moodle are based on social constructionist pedagogy called Moodle philosophy (Dougiamas, 2022). This is also applied to the design and development of other LMS such as Claroline, Sakai, Dekeos, etc.

The features of an LMSs are designed to support collaborative learning. Features such as forums, wikis, glossaries, databases, messaging and chats allow students to control common content to support group collaboration.

The learning platforms also allow students to express and share knowledge. For example, wikis are features built into an LMS to be used for group work and negotiation. Glossaries define terms that appear throughout the course.

Students can learn by observing the activities of each other on an LMS. Students can view the activities of each other on the course using the participants' feature. Students can also see everyone who is online at the same time.

Students can understand each other's learning context to enable instructors to teach in a more transformative way by understanding the context of each learner

An instructor can teach in a more transformational way learners provide information about their background and express themselves publicly. Learners can modify their user profiles on an LMS and provide information about their backgrounds. Through individual blog features that appear in most LMS, learners can share knowledge in public for others to provide feedback or comments.

LMS provides learning environments accommodating adaptability and flexibility to encourage response to learners' needs. In most LMS, a teacher can visit a course page and structure learning activities to accommodate learners. Furthermore, the grade book feature is automatically generated and can show learners' activities in a course.

### **2.1.5 The Weaknesses of an LMS in the Context of Information Processing Theory**

LMSs provide a standard framework for organizing and systematically delivering learning materials. Even though these LMSs provide a variety of features that are used to enhance learning such as quizzes, forums, chats, assignments, and wikis, they largely lack features for identifying

learning preferences and filtering contents to match the cognitive behavior of learners. The development of LMS though is based on social constructionist pedagogy, its focus is to apply general teaching methods without considering the individual needs of a learner. This implies that instructors only apply teaching preferences that may not necessarily address the individual needs of a learner. Instead, a learner must fit into the instructional method of an instructor.

Learners with different cognitive preferences act differently in an online course and use different features such as examples, and exercises with different frequencies, and visit learning content in a different sequence (Graf and Kinshuk, 2013). A study by Usman *et al.* (2021) established that learners have difficulties in learning if their behavior is not supported by an environment and that implementing effective learning models can improve the metacognitive skills and retention ability of learners.

As stated by Felder (1988) an instructor should provide a course with many features supporting different learning behavior. Therefore an effective LMS must incorporate all these essential features.

As a response to address the identified weaknesses of an LMS, this thesis proposes an automatic modeling method, for continuously monitoring and analyzing learner navigation patterns in an LMS. The study demonstrates modeling methods for identifying indicators of psychological theories and using the data to estimate learner behavior. A rule-based prototype is designed, implemented, and evaluated with the students in a virtual classroom environment. The prototype can be integrated with an LMS to identify the cognitive preferences of the learners.

## **2.2 Learning Style Models**

Learning is defined as the process by which an individual acquires knowledge, behaviors, skills, values, attitudes, and preferences (Gross, 2012). Individuals have different manners of approaching learning tasks and strategies.

Different scholars have varying definitions for the term learning style. According to Blakemore, McCray, and Coker (1984), learning style is the way a learner observes, interacts with, and makes a response to learning content. Cassidy, (2004) states that it is how an individual approaches a learning situation.

Even though learning styles models are varied, each theory argues that people learn in different ways and instructions can be tailored to fit-individual preferences. For instance, some people learn best by watching, some by listening, and some by movement.

Learning style, as such, defines a learner's concentration, processing, absorption, and retention of new and challenging information. Individuals show these characteristics differently. Instructors need to know how to trigger and maintain an individual's concentration to produce long-term memory and retention. Using a comprehensive model that identifies an individual's strengths and preferences considering physiological, sociological, psychological, emotional, and environmental elements is useful for revealing these natural tendencies and styles (Dunn and Burke, 2006). Some may want explanations for grammatical rules; others may not need explanations. Some may feel writing down words or sentences helps them to remember them, while others may find they remember things better if they are associated with pictures (Ababneh, 2015).

Several cognitive and learning style models for assessing students' learning differences exist in the literature. Here, we will take a look at some of the most commonly used LS models as outlined below.

### **2.2.1 Felder-Silverman Learning Style Model**

In the Felder-Silverman learning style model, learners are grouped as active-reflective, sensing-intuitive, sequential-global, and visual-verbal (Felder, 1988).

According to Felder(1988), active learners work best by putting the knowledge acquired into practice and trying things out. They like communicating with others and working in groups to discuss the materials learned. In contrast, reflective learners think and imitate the materials learned. They work alone and sometimes in small groups.

Sensing learners read details and actual materials. They tackle problems using standard methods and are keen on details. They are more practical, like relating the knowledge acquired to the real world. In contrast, intuitive learners like reading theories and their underlying connotations. They can identify possibilities and their relationships. They are more innovative and creative.

Visual learners like content that they can see such as images and videos. Verbal learners prefer written and spoken content.

Sequential learners navigate content step by step. They follow linear learning progress. They follow systematic steps to find solutions.

Global learners are holistic thinkers. They learn in large leaps. They read by skipping contents. They absorb learning materials randomly without being keen on connections but suddenly get a whole picture after navigating the entire content.

There is an Index of Learning Style (ILS) questionnaire for collecting data based on FSLSM. The questionnaire consists of 44 questions with two options categorized into four dimensions representing active - reflective, visual-verbal, sensing-intuitive, and sequential-global.

### **2.2.2 Honey and Mumford Learning Style Model**

This learning style model classifies learners as activists, reflectors, theorists, and pragmatists (Honey and Mumford, 1982). Activist learners learn by doing. They learn best through brainstorming, problem-solving, and group discussions. Theorists learn best through learning activities such as models, statistics, stories, quotes, and background information. Pragmatists learn best through the application of learning in reality, case studies, problem-solving, and discussions. Reflectors learn best by watching and contemplating what happened. They look at alternate points of view, gather information, and work toward a suitable conclusion. They learn best through paired discussions, self-analysis questionnaires, personality questionnaires, time out, observing activities, feedback from others, coaching, and interviews.

Honey and Mumford's Learning Styles Questionnaire (LSQ) was designed by Peter Honey and Alan Mumford. The questionnaire consists of 80 questions representing activist, theorist; pragmatist, and reflector learning styles.

### **2.2.3 Myers Briggs Type Indicator (MBTI)**

In the Myers-Briggs Type Indicator learning style model, learners are classified as extraversion - introversion, sensing - intuition, thinking - feeling, judging - perceiving (Myers, 1995).

Extroverts try things out. They are outgoing and concentrate on the outside world of thoughts.

Introverts are inward and thinking, concentrating on the inside world of thoughts.

Sensors are hands-on and like details. They concentrate on facts and procedures. Intuitionists are creative and concept-oriented. They focus on meanings and possibilities.

Thinkers make decisions based on logic and rules. Feelers make decisions on personal and humanitarian thoughts.

Judgers set and follow the agenda. They seek conclusions even with incomplete information. The perceiver adjusts to changing circumstances and postpones concluding to obtain more information.

Myers Briggs Type Indicator instrument was constructed by Katharine Cook Briggs and Isabel Briggs Myers. The questionnaire consists of questions categorized into four dimensions representing introversion - extraversion, sensing - intuition, thinking - feeling, judging - perceiving learning styles. A four-letter test result is produced by taking one letter from each category to infer a learning style.

#### **2.2.4 VARK Learning Style Model**

VARK is an acronym used to describe learner characteristics (Fleming, 2014). The characteristics are visual, aural, read/write, and kinesthetic. Visual learners like illustrations and visual content. Auditory learners prefer verbal lessons. They prefer learning by talking and listening. Read/write learners prefer information displayed as text. They prefer content in form of text-based inputs and output. They learn best by reading and writing. Kinesthetic learners prefer a hands-on approach to learning. They prefer to learn by moving, touching, and doing.

VARK questionnaire was designed by Fleming. The questionnaire consists of 16 questions with four options representing visual, aural, read/write, and kinesthetic.



### **2.2.5 Kolb's Experimental Model**

In Kolb's experimental model, learning is classified as concrete experience - abstract conceptualization, active experimentation - reflective observation (Kolb, 2015). Concrete experience - abstract conceptualization describes how learners take the information that is as an approach to grasping experience. Active experimentation - reflective observation describes how learners process information. Reflective observation and active experimentation are related to the transformation of experience. Kolb identifies four types of learners in this classification scheme. The classification schemes are concrete experience - active experiment, abstract conceptualization - active experiment, concrete experience - reflective observation, and abstract conceptualization - reflective observation. Learners with strong hands-on practical activities e.g. physical therapists are classified as having concrete experience and active experiments. Learners with strong hands-on practical application of theories e.g. engineers are classified as those possessing abstract conceptualization and active experiments. Learners who are strong in imaginative ability and discussions e.g. social workers are in the concrete experience and reflective observation category group. Learners who have strong inductive reasoning and the creation of theories e.g. philosophers constitute the abstract conceptualization and reflective observation group.

Kolb's Learning Style Inventory (LSI) was designed by Kolb. The questionnaire consists of 10 questions representing accommodating, converging, diverging, and assimilating learning styles.

### **2.2.6 Dunn and Dunn's Learning Style Model**

Dunn and Dunn's learning style model consider five different actors. They include environment, emotion, society, physiology, and psychology (Dunn, 1990). The Dunn and Dunn Model suggests that students are affected by these factors. Students respond differently based on various factors in a learning environment e.g. listening to background music, others quiet environment; some study

under bright lighting, others prefer dim; some study in a warm environment, others cool. To ensure higher retention rates and greater achievements, teachers need to identify and address the appropriate environmental preferences. Some students feel uncomfortable and tense learning under pressure and have difficulty concentrating in teacher-dominated and authoritative classrooms. Students who prefer learning either alone or with their peers may feel uncomfortable and tense concentrating in teacher-dominated classrooms. Some students have different ways of interacting with the learning environments. Some students are auditory, visual, tactual, and kinesthetic. To achieve academic gains, teachers need to match students' time preferences with learning and testing schedules.

Dunn and Dunn's learning style instrument, *The Learning Style: The Clue to You! (LS: CY!)* was designed by Rita Dunn and Kenneth Dunn. The instrument consists of 69 questions categorized into four dimensions representing the environment, emotions, social, physiology, and psychology preferences that each student has for learning.

### **2.2.7 The Herrmann Brain Dominance Instrument (HBDI)**

This instrument identifies four different modes of thinking. The modes of thinking identified are based on analytical, sequential, interpersonal, and imaginative behavior (Herrmann, 1976). Learners that prefer activities to help them understand how things work are analytical thinkers. Learners who prefer a step-by-step approach to solve problems, organizing and implementing activities are sequential thinkers. Learners who prefer to listen to and express ideas, work in groups, and interact with others are interpersonal thinkers. Learners who think and look at the big picture, take initiative, challenge assumptions and engage in creative problem-solving and long-term thinking are imaginative thinkers.

The Herrmann Brain Dominance instrument was designed by William "Ned" Herrmann. The instrument consists of 116 questions categorized into four dimensions representing analytical, sequential, interpersonal, and imaginative thinking.

Understanding the concepts of learning style models helps a researcher conceptualize the relationships between the learner navigation patterns in an LMS and the psychological characteristics under investigation in learner behavior modeling research (Moraes *et al.*, 2020; Correa, Monsalve, and Tabares, 2020). This can help a researcher to infer learning styles from patterns of learner interactions with an LMS.

### **2.3 Cognitive Traits Important for Learning**

The cognitive traits profile an individual according to the cognitive abilities which are identified as cognitive traits, working memory capacity, inductive reasoning ability, information processing speed, and associative learning (Lin and Kinshuk, 2004; Lin, 2007; Graf and Kinshuk, 2008). According to Deary *et al.* (2004), the cognitive traits are still valid due to the more persistent nature of the cognitive behavior of learners which changes according to the new learning environment. This confirms that cognitive traits are still valid in research concerning learner behavior modeling to date. This section discusses the four cognitive traits in detail.

#### **2.3.1 Working Memory Capacity**

The ability to briefly keep a limited amount of information in the mind is determined by working memory capacity (Miller, 1956). To minimize cognitive overload in the mind of a learner, learning materials must match an individual's working memory capacity. Cognitive load is reduced by decreasing the number but increasing the relevance of paths in a course for learners with low working memory capacity. Such a learner requires more concrete content. The number of media

resources should also be increased. A learner with high working memory requires less relevant paths.

Working memory capacity has traditionally been measured using the operation span tasks (Turner and Engle, 1989). The Operation Span (OSPAN), a measurement tool for working memory capacity is an online version developed in PsychoPy2 (v1.83.04) (Lurquin and Michaelson, 2017). In the web span, there are 60 arithmetic operations and 30 words that a student answers to test the working memory capacity. Participants are subjected to recall in their correct order sequentially presented words while solving simple math equations concurrently.

### **2.3.2 Inductive Reasoning Ability**

Inductive reasoning skills are related to the extent to which a student can formulate ideas from cases by making broad generalizations from facts to come up with a general conclusion. Conclusions are drawn from data i.e. from specific facts to general by making many observations, discerning patterns, making generalizations, and inferring explanations (Bradford, 2017). Learners with low inductive reasoning require opportunities for observation. Therefore, these learners should be provided with well-structured and concrete information with many paths. Learners with high inductive reasoning skills require information with fewer paths and complexity to enable them to grasp concepts quickly. Such information can be presented in abstract form

Inductive reasoning has traditionally been measured using Wason 2-4-6 hypothesis Rule Discovery task (Wason, 1960). This is a hypothesis-testing tool designed for testing logical reasoning with tests to demonstrate the processes. The tool is used for testing the ability of an individual to find out relevant attributes shared in the stimuli by comparing odd, even, ascending, and descending.

### **2.3.3 Associative Learning Ability**

The specific experiences associate psychological links between concepts, events, and mental states (Klein, 2012). A learner's ability to formulate a relationship between the current and the original state is defined by associative learning ability. The compactness of the related state shows the ability to evoke sensory images that can function as connections of associative learning and memory (Paivio, 1969).

Thus associative learning connects the current to existing knowledge. An instructor needs to enhance the revisiting of learned materials and clearly illustrate the relationships of the new to previous concepts to support an associative learner.

An instructor needs to provide a high amount of information, different media resources, and many relevant paths to help a learner with low associative learning ability. Such a learner needs well-structured information to make the linkage between new and existing concepts easier.

An instructor needs to provide a learner with high associative learning skills with less structured information to allow free navigation and enhance learning speed.

Web-dal is a web-based test for associative learning (Lin, Kinshuk, and Graf, 2007). A subject is first presented with a new concept and learns and writes related concepts in one minute.

### **2.3.4 Information Processing Speed**

Information processing speed is articulated based on response time, the time it takes to finish a series of tasks, or the number of items responded to correctly within a given time (Sweet, 2011).

Information processing speed measures the effectiveness of the cognitive operation. Information processing speed is assessed using timed tests that challenge cognitive operations. A learner's ability to acquire and apply the information correctly is determined by information processing

speed. Instructors should only provide important points where a learner has low information processing speed. This means decreasing the number of paths and information but increasing the number of relevant paths. Such information structure helps to increase the learning speed of such individuals. In contrast, learners with high information processing speed are better served by enhancing the information by providing a high amount of information and paths. The number of paths and amount of information can be increased for these learners.

Mental speed tests are used by psychologists to test information processing speed (Perina et al., 2021). The mental speed tests measured how quickly a student can process information and make decisions based on the available information. For each timed true-false question, a graphical word/image pair is provided. A student is expected to follow specific instructions to attempt the questions. A student is expected to click the True button if the pair matches and false in case of the opposite answer. The answer is reversed if the word "opposite" appeared. The number of objects and boxes increases as the student progresses with the questions. Performance is indicated by the number of correct answers.

Understanding the concepts of the cognitive traits model helps a researcher conceptualize the relationships between the learner navigation patterns in an LMS and the cognitive behavior under investigation in learner behavior modeling research (Moraes *et al.*, 2020; Correa, Monsalve, and Tabares, 2020). This can help a researcher to infer cognitive behavior from patterns of learner interactions with an LMS.

## 2.4 Learner Modeling

Individualized online tutoring demands an LMS that keeps a profile with the personal preferences of a learner. This is accomplished through learner modeling.

A model is a human-computer interaction technique for creating and adjusting a theoretical understanding of a user by customizing and adapting a system to user needs through internal representations (Fischer, 2001)

A user model collects and categorizes personal data about a given user. The model captures certain characteristics about an individual to create a profile. User modeling is the process of creating a profile. Data included in the user profile may include personal information, interests, skills, knowledge, and preferences (Piao and Breslin, 2018).

Different design patterns can be used to create user model profiles namely static, stereotype, and dynamic (Johnson and Taatgen, 2005).

Data in the static models are not modified once captured in the user profile. An example in the literature is a study conducted by Abdelhadi, Ibrahim, and Nurunnabi (2019) to determine students' learning preferences using an LS questionnaire issued to students. The students were grouped into clusters based on the similarities of their preferences according to the information provided in the LS questionnaire.

Dynamic models, on the other hand, allow modifications of the profile to accommodate the current goals and needs of the user. An example in the literature is a student model for identifying learning styles and affective states presented by Khan *et al.* (2019). In this model, the learning preferences and behavior are gathered when students interact with the course. From this, learning styles and affective states are automatically generated.

Stereotype models use demographic statistics to create a user profile. Predictions about a user can be made based on the assumption that those sharing demographic profiles have similar characteristics. An example is a study conducted by Harrak, Bouchet, and Luengo (2019) to identify the relationship between students' questions and their profiles based on existing data. Learner navigation on contents and online assessment attempts were analyzed using a coding scheme. Based on the learner profile created, a model was developed to predict future learner profiles in the next iterations of the course.

#### **2.4.1 Gathering Data for Modeling**

There are different methods for gathering data for creating a user profile.

A user can be asked to provide specific facts about himself or herself when accessing the system (Johnson and Taatgen, 2005). Feng, Iriarte, and Valencia (2020), for example, administered the Honey-Alonso Learning Styles Questionnaire (CHAEA) and an adaptation of the Strategy Inventory for Language Learning (SILL) questionnaire to students to examine the relationships between learning styles, learning strategies, and academic performance of Chinese students. The students filled out the questionnaires to provide facts about their learning styles and learning strategies. While this approach is a convenient way to quickly collect data, it is limited by the fact that the model cannot automatically adjust to match the user preferences. The method also depends on the user's willingness to disclose full information. The user profile is therefore unlikely to contain up-to-date information.

Johnson and Taatgen (2005) indicated that users' preferences can be gathered by observation and interpretation of behavior during their interactions with the system. In a related study, Boticki, Akçapınar, and Ogata (2019) analyzed log data on e-book usage to generate user models describing



students' user engagement and user reading style variables. This method is limited by the fact that ascertaining learning preferences by interpreting the behavior might appear unclear to a user who might not understand the system behavior. Furthermore, the system must gather a certain amount of data to accurately predict a user's preference (Montaner, López, and De La Rosa, 2003). But the provision of an automatically adjusting user model allows it to predict user preference with good accuracy once adequate user data is gathered.

The advantages of both methods can be achieved by applying a hybrid approach when gathering modeling data. A user can be asked to provide specific facts and provide feedback about himself or herself. The interactions and behavior during interactions with the system can also be observed. The hybrid information gathered is used to automatically modify the model to accommodate user preferences (Montaner, López, and De La Rosa, 2003).

#### **2.4.2 Learner Adaptation**

Using an analytical algorithm relevant features of information can be presented according to a user's interests and goals as captured in the user model. This is done by displaying only relevant features and hiding the rest. This can be implemented using adaptable and adaptive systems (Fischer, 2001).

In an adaptable system, a user is provided with corresponding options describing the preferences. A user can manually select the relevant option to display information matching the behavior. For example in Chang et al. (2016), a learner fills out an LS questionnaire during the first login on an LMS to determine learning preference. The system then provides appropriate teaching content, according to the questionnaire results.

In an adaptive system, a rule-based filtering algorithm interprets information stored in the user profile to display information matching the behavior (Johnson and Taatgen, 2005).

Another version of the adaptive system is using a collaborative filtering algorithm (Montaner, López, and De La Rosa, 2003). Here information stored about a user is compared with others in the system. The system assumes that where the model about a certain user is unknown, the person is likely to share similar characteristics with those with a similar profile.

## **2.5 Learner Behavior Modeling in an LMS**

As discussed in the previous sections, LMS unlike face-to-face teaching does not have the features for tailoring teaching to support the behavior of each student. Information processing theories suggest that individuals have different styles of processing information in the mind. Thus the instructional method must match an individual behavior.

In this section, the studies that apply learner modeling in the context of computer-based learning are examined. It looks at learner modeling data and profile characteristics that can be extracted from an LMS, the process of associating behavior patterns with psychological theories e.g. LS and CT, and also undertake a review of the methods of identifying learner behavior applied in related studies.

### **2.5.1 Learner behavior data that can be extracted from an LMS**

Collecting data for investigating learner behavior in an LMS should focus on providing a coherent environment for individual learning preferences and interests (Moraes *et al.*, 2020). This demands understanding the relationship between the learning object visited by a student and learning behavior (Correa, Monsalve, and Tabares, 2020).

Learning Objects (LO) are instructional resources and activities provided by an LMS. Examples of LO provided by LMS such as Moodle LMS (*Moodle*, 2018) and Claroline (*Claroline*, 2017) are:

- Lessons that allow instructors to create a sequence of interactive exercises
- An assignment that allows a student to submit work for grading
- quizzes allowing an instructor to create online tests
- the forum which allows students to hold discussions on a specific topic
- the file that allows instructors to upload learning handouts as word-processed or slideshow presentations.

An LMS collects data when a student interacts with the learning objects. LO can motivate the learning process and promote meaningful teaching (Nafea, Siewe, and He, 2019; Sensus, Hasani, and Bagustari, 2020).

In a study to identify learning styles and affective states in web-based Learning Management Systems (LMSs), Khan *et al.* (2019) identified LO in each behavior investigated. In this study, data for identifying the psychological behavior under investigation were collected. The study applied commonly used learning objects such as exercises, content objects, self-assessment tests, discussion forums, illustrations, assignments, outlines, and students' navigation behavior within the course. The data collected from each learning object were based on their pertinence for the psychological behavior being investigated. For example, the patterns of visits and stay were selected against the content objects which describe the total visits and the time spent by learners over the contents. The number of visits, posts, and replies were selected against the discussion forum objects that describe the participation of a learner in online discussions. The patterns of

assignment revisions and stay were selected against the assignment objects, which describe the interactions with the assignment object.

Renato (2020) developed an adaptive virtual learning platform for personalizing the course content design and user interfaces to match the behavior of learners. In this study, the number of discussion posts was used against the object forum. The number of visits and time spent were used against the learning objects concrete materials, examples definitions, theories, textual materials, and illustrations.

Assis *et al.*(2022) in a systematic review to discover how learning behavior proposed by psychological theories relates to LO in an LMS reported that learning behavior such as LS was related to several LO indicating that education and pedagogical theories were taken into account to provide appropriate learning resources for each student. The study, however, concluded that due to inconsistencies in the results, more in-depth research was needed to allow a more accurate understanding of the relationships between learning objects and learning behavior.

This thesis thus enriches the previous work on the relationship between learning objects and psychological theories by identifying the patterns of learner interactions with the contents in an LMS. The identified patterns can be used to infer LS and CT from an LMS.

### **2.5.2 Associating LMS Behavioral Patterns with LS and CT models**

In their efforts to provide an automatic student modeling approach for an LMS, Graf, Viola, and Kinshuk, (2007) used patterns extracted from features commonly used by instructors and are implemented in most LMS. The LMS features considered in the study included: outlines, detailed content, quizzes or exercises, self-assessments, examples, and discussion forums. Outlines provide a summary of the contents covered in a chapter. Detailed contents are topics with facts and

concepts discussed in detail. Self-assessment tests are learning activities that students use to check their acquired knowledge. *Exercises* are test examination questions generated from the contents covered for students' practice. Examples are cases related to the concepts discussed in the course. Discussion forums are questions assigned for students to deliberate in groups to learn from each other. To identify LS patterns, the number of times a student visits a learning object is considered. The time a student spends viewing a relevant learning object is also considered for analysis. Examples of such learning objects are course outlines, topics, sub-topics with detailed notes and facts, and content objects with examples. The content objects with self-assessment questions and quizzes are also considered. For example, the total number of questions a student answers and the time spent attempting tests are considered for analysis. The number of revisions made on tests is also considered for analysis. The time a student spends reviewing test results is considered for analysis. The performance on questions dealing with a specific type of content for instance concepts, details, overviews, graphics, and text is considered for analysis. Student participation in discussions is also considered for analysis. The number of discussion posts and replies made by a student is considered a pattern for analysis. The time a student spends reading forum posts is equally considered a pattern for analysis.

### **Relating Students' Behavior with LS Preferences**

Studies conducted to identify LS based on Felder-Silverman Learning Style Model (FSLSM) argue that the browsing patterns of behavior on an LMS are assumed to be related to each learning style dimension (Graf and Liu, 2009; Graf and Kinshuk, 2013; Khan *et al.*, 2019). These works detect learning styles by analyzing the behavior of students in an LMS on the use of learning content. This section outlines the techniques for inferring LS from content browsing behavior and matching with relevant dimensions of FSLSM.

### ***Active–Reflective***

According to FSLSM; active learners discuss, apply and explain what is learned to process the information. Reflective learners learn by thinking about the materials learned. A student who frequently posts on a discussion forum attempts a high number of exercises and self-assessment tests is an active learner. A student who posts less frequently on the discussion forum makes more visits to learning materials, and spends time viewing end-of-unit tests is an indication of reflective LS (Renato, 2020; Assis *et al.*, 2022; Khan *et al.*, 2019).

### ***Sensing – Intuitive***

Sensing learners concentrate on concrete materials such as examples. They focus on facts and data. They tend to solve problems following well-established procedures. They work slowly, are patient, and are cautious with details. They perform well using repetition as a learning technique. Intuitive learners prefer challenging tasks and dislike details and repetitions. Such learners also work at a fast pace. A student who frequently visits content with examples is a hint for a sensing LS. As well, a student taking more time to revise before submitting exercises and eventually getting a reasonable score is a sensing learner. Where there are frequent content visits, and a learner spends less time on content with examples is an intuitive learner. Making fewer or no revisions on self-assessment tests and eventually scoring a modest mark on the first attempt shows disliking repetitions demonstrates an intuitive LS (Khan *et al.*, 2019; Assis *et al.*, 2022)

### ***Visual–Verbal***

Visual learners prefer learning materials with images and videos. Those with verbal learning habits prefer content in the form of words, whether spoken or written. Therefore, frequently visiting and spending more time on content with graphics and scoring moderate grades on questions with graphics and video are visual learners. On the other hand, frequent visits to content with words is

a sign of verbal LS. Such learners have the habit of discussing and communicating with others and thus are identified with frequent posts and replies on the discussion forum (Assis *et al.*, 2022).

### ***Sequential – Global***

Sequential learners solve problems following a logical order by navigating through the learning materials linearly. Global learners prefer getting a full synopsis of the contents without getting into many details. Patterns related to a sequential learner can be gathered by observing the order in which a student is following contents in an LMS. Thus few navigation skips indicate a sequential LS. On the other hand, the number of content pages skipped, and visits to content overviews and outlines are characteristics of a global LS (Assis *et al.*, 2022; Al-Omari, Carter, and Chiclana, 2016).

### **Relating Students' Behavior with CT Preferences**

When learners browse the contents of an LMS, their activities are recorded in the system interaction logs. The navigation patterns can be retrieved from an LMS database and used to infer CT based on the behavior described in CTM. This section discusses human observable behavior in an LMS which can be used as an indicator of CT.

### ***Working Memory Capacity***

In an LMS, content navigation patterns, the ability to resist interference, accessing external resources not provided in the contents, simultaneous content access, and performance on end-of-unit tests are signs of high working memory capacity (Graf and Lin, 2008). Linear navigation of contents demonstrates that a student can concentrate on a learning task thus indicating high working memory. Nonlinear or reverse navigation means low working memory capacity. The frequency of visits to external contents and performance in a test related to external learning

resources is an indicator of working memory. Frequent visits to external learning resources and good performance on end-of-unit tests are signs of high working memory (El-Bishouty *et al.*, 2015).

### ***Inductive Reasoning Ability***

The ability to filter data relevance, pattern matching, generalization, background knowledge, utilize local hypothesis strategy, learn from analogy, and classification ability are signs of inductive reasoning ability (Sabine, Graf, and Kinshuk, 2008). In an LMS, performance on exercises following examples or case studies is an indicator of inductive reasoning. Performance on content that contains a parallel concept i.e. an alternative point of view of a concept and to what extent a student can complete a set of procedures as part of learning indicates inductive reasoning ability.

### ***Associative Learning Ability***

The ability to relate concepts is an indication of associative learning ability (Sabine, Graf, and Kinshuk, 2008). In an LMS, visited hyperlinks to view relevant contents outside the curriculum and linearity of navigation of contents are indicators of associative learning (Lin, Kinshuk, and Graf, 2007).

### ***Information Processing Speed***

Information processing speed checks how fast the learner acquires information and applies the knowledge to make correct decisions (Sabine, Graf, and Kinshuk, 2008). Some learners have slow reading speeds while others are fast readers. In an LMS, the time taken in viewing contents can indicate information processing speed. The time taken answering questions and the grade attained on the test is also can indicate information processing speed.



## **2.6 Estimating LS and CT from Behavior in an LMS**

Section 2.5.2 highlighted relevant patterns for inferring learner behavior from LMS logs. These human observable indicators in an LMS can be used to infer LS and CT. Data related to a particular learner preference can be used to estimate the hints for a particular psychological behavior e.g. LS and CT. This section discusses methods for estimating LS and CT using records of learner interactions with an LMS.

### **2.6.1 Machine Learning Methods**

Mitchell (1997) defines learning, in artificial intelligence, to include any computer program that improves its performance at some task through experience. A computer program is said to learn from experience  $E$  concerning some tasks in  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $O$ , improves with experience  $E$ . Machine learning algorithms can build a model based on training data to make predictions without being programmed. Bishop (2006) suggested that the machine learning algorithm can be expressed as a function  $y(x)$  which takes  $x$  as input and that generates an output vector  $y$ , encoded in the same way as the target vectors. The precise form of the function  $y(x)$  is determined during the training phase, also known as the learning phase, based on the training data. Once the model is trained it can then determine the identity of the new output comprising a test set. The ability to categorize correctly new examples is known as generalization. The original input variables are preprocessed to transform them into some new space of variables. This stage is called feature extraction.

Machine learning methods have been used to investigate learner behavior as can be seen in several studies cited below. In most of these studies, both supervised and unsupervised learning methods are used to estimate learner behavior in an LMS.

### ***Bayesian networks analysis***

Using Bayesian networks, data extracted from an LMS can be analyzed to predict the probability of a learner having a specific learning preference given the evidence. One such study was conducted by García, Schiaffino, and Amandi (2008) to predict LS based on FLSM from LMS data using conditional probabilities as:

$p(V_i=v_i/E=e)$  i.e. some variable  $V_i$  has value  $v_i$  given the evidence

$p(LS=LS1/pattern1,pattern2..)$  e.g.  $p(LS=Sequential/content\ Access, results)$

In another related study, Rajper *et al.* (2016) analyzed LMS data to produce the Conditional Probability Tables (CPT) for each learning style using each attribute. The results were compared with those gathered from - Kolb's Learning Style Inventory (KLSI) questionnaire issued in class reporting 50% and 75% accuracy for each learning style dimension.

### ***Analysis using intelligent agents***

Learner behavior modeling in an LMS has also been addressed by analyzing data using intelligent agents. Kaplana and Haenlein (2019) argued that an intelligent agent is a system that can correctly interpret external data, by learning from it and using the learning experience to achieve specific goals and tasks through flexible adaptation. Chang *et al.* (2016) developed an agent-based model for estimating LS using data collected from an LMS and allocating relevant learning resources. Based on the adaptive learning strategy, a learner fills out the ILS questionnaire during the first login to ascertain LS. The user's learning styles are determined based on the ILS questionnaire results. The learning materials are provided based on the learning styles detected according to an individual. An experiment conducted to evaluate the effectiveness of the agent-based model showed that 80.65% were active, 88.71% visual, 69.31% sense, and 62.9% global learners.

In another related study, Hasibuan et al. (2016) proposed a model based on an intelligent agent. The model does an initialization, learns, matches, and recommends the learning styles of an individual. However, this study was not evaluated against any experiment with students. In another study, Kamsa, Elouahbi, and El Khoukhi (2018) used three intelligent agents to customize and adapt the learning time focusing on an individual concentration time. The model was evaluated by administering tests and questionnaires to 50 students reporting dependence between concentration time and success of the learners with 70% satisfaction. The study however needed to be evaluated by more students in a real learning environment.

### ***Decision Tree Analysis***

Using browsing patterns collected from an LMS as training input and data collected using a learning style questionnaire as output; Liyanage, Gunawardena, and Hirakawa (2016) used the decision tree algorithm to estimate learner behavior. The study applied a classification technique to predict LS by analyzing log data extracted from Moodle LMS based on the J-48 algorithm of the Waikato environment for knowledge analysis (WEKA). Comparison of predicted against actual results gathered from ILS indicated accuracy rates of 65.26%, 80.00%, 90.00%, and 81.25% for active-reflective, sensing-intuitive, sequential-global, and visual-verbal respectively. This study despite the impressive results was specific to Moodle LMS and needs to be evaluated in other LMS.

### ***Neural Network Analysis***

Using neural network analysis, records of students' interactions with an LMS are fed into the network as input. The specific learning attributes can be used as an output to be predicted by the network. An example of using artificial neural networks (ANN) to predict specific learner attributes was conducted by Aditya and Babita (2018) to classify student learning attributes. The

attributes classified in the study were anxiety, personality, learning style, cognitive style, grades from the previous semester, motivation, study level, and student prior knowledge. The study classified the learning attributes using a feed-forward neural network and a Least Square-Support Vector Machine.

Yu, Wu, and Liu (2019) also analyzed records of learning materials access, video access, and assessment grades to predict learning outcomes. The study reported the results of 0.950837989 prediction accuracy.

In another related study, Ahmad and Shahzadi (2018) used records of prior grades, an environment where students live, study habits, learning skills, ability to work hard, and academic interactivity as input to predict the likelihood of being at risk or not reporting. The study results reported between 95% and 85% accuracy.

Gautam et al. (2018) used an artificial neural network classifier to categorize speech acts with 0.796, 0.795, and 0.781, precision, recall, and f-score respectively.

Just like the decision tree analysis mentioned above, records of students' interactions with the system can be used as input into the network, and data collected using manual psychometric tools as output for prediction. Begum and David (2017) conducted a study to analyze data collected from Kolb's questionnaire using the Fuzzy Min-Max Cascade Correlation Neural Network (FMMCasCorNN) to predict students' learning styles. An evaluation experiment conducted on 180 students showed classification accuracy of 0.918, 0.926, 0.922 precision, recall, and f-measure. Sweta and Lal, (2016) as well, in their study predicted LS from LMS logs using fuzzy logic and neural network. The results were compared against data collected from the Index of Learning Style (ILS) issued to 60 students taking a course on C++ programming. Evaluation results

were 71.25%, 76.12%, 75.25%, and 77.35% precision for active-reflective, sensing-intuitive, visual-verbal, and sequential-global. Using input data from previous studies and the ILS questionnaire issued to 75 students as output, Bernard et al. (2015) conducted a study to predict learning styles. The study reported precision results of 0.802, 0.741, 0.727, and 0.825 active-reflective, visual-verbal, sensing-intuitive, and sequential-global precision.

### ***Analysis Using Rules***

Using rules, researchers have proposed methods for estimating learner behavior. An example of such a study is by Latham *et al.* (2012), where learning styles are predicted using logic rules. The rules are formulated from the data captured from an Index of learning style questionnaire. The model uses IF-THEN as illustrated below:

IF the student has shown an image/diagram

AND the student gives the correct answer

THEN increase VISUAL;

### ***Cluster Analysis***

Clustering analysis measures the distance between the objects to place them in groups with similar features. The common methods measure minimum, maximum, mean, and average distances between objects (Han, Kamber, and Pei, 2012).

Abdelhadi, Ibrahim, and Nurunnabi (2019) applied group technology to classify students according to their LS preferences. ILS questionnaires were issued to 44 students to determine their LS preferences. Multivariate analysis was used to group students into clusters based on the similarities of their LS. The common LS preference reported the highest similarity of 93.5%. The

study suggested that instructors can use the developed groups to apply teaching styles matching the behavior of their students.

Poellhuber and Roy (2019) classified learner profiles based on their behavior in an open-source LMS Sakai to study the survival in the massive open online courses (MOOC). The data relating to lecture slides and streamed videos viewed, reading materials downloaded, the quizzes attempted and submitted, forum questions asked and answers contributed were analyzed using cluster analysis procedure to classify learners as browsers, self-assessors, serious readers, active-independent, and active social. The study reported that the five profiles reported different levels of engagement. It was also observed that a significant percentage of the learners belonging to the less active profiles did not drop out of the MOOC.

Lee and Natriello(2018) conducted a study to gain insights into patterns of user interaction in an online video discussion platform called Vialogues. The data relating to video clickstream were analyzed using a hierarchical clustering analysis to classify learners as video watchers with no discussion, opinion seekers and active repliers with little or no video-watching activity, users who watch and discuss videos and, users focused on viewing and creating metadata. The study concluded that the findings could be used as a reference for designing video discussion tools for teaching and learning.

These studies were reviewed to demonstrate how machine learning algorithms learn and create predictive models by moving from specific examples to a more general description of the data. For example, when grouping students according to their browsing habits as demonstrated by Poellhuber and Roy (2019), or classifying student learning attributes using a neural network as presented by Aditya and Babita (2018), errors can always be produced during the training process.

The concern with machine learning methods is the effect an incorrect and correct classification will have on identifying learner behavior. When a student is placed in a wrong learning behavior category because of a classification error, an instructor can subject such a learner to an instructional method not matching the learning preferences. Evidence on whether information generated through machine learning models improves learning in practice is still scarce (Ferguson and Clow, 2017; Guzmán-Valenzuela *et al.*, 2021; Knight, Gibson, and Shibani, 2020; Rets *et al.*, 2021; Wilson *et al.*, 2017) and still requires further investigation.

Another shortcoming of machine-learning methods is that they produce black box models without showing transparency into their internal working mechanisms (Adadi and Berrada, 2018). This makes it difficult for users to understand how the machine learning models make predictions and what aspects of the behavior are driving the prediction. This lack of explanation can lower the application and trust of such models by users (Baneres *et al.*, 2021).

To address the weaknesses pointed out, this thesis demonstrates how artificial neural networks and cluster algorithms can analyze the data set and classify the students according to their learning behavior. To increase the prediction accuracy of the machine learning models, the data set used in the study was labeled to match the psychological variables under investigation. The data set in the study was normalized by converting numeric values to a scale of  $[0,1]$  to change the range and have a normal distribution. Training the neural network was done by gradually increasing the hidden layers of the network and recording the results. This reduced network overfitting and minimized the model complexity.

To increase classification accuracy with cluster analysis, the four algorithms, commonly used for analysis hierarchical, k-means PAM, and Diana was used. The analysis was repeated using each algorithm by increasing the number of cluster members from 1 to 10 and recording the results.

Fränti and Sieranoja (2019), reiterated that repeating the algorithm reduces error further by at least 10% to ensure more accurate classification and reliable modeling information.

### **2.6.2 Estimation by Analyzing Browsing Patterns Using Rules**

This technique uses the actions of learners on an LMS to get hints about their learning preferences and then applies a simple rule-based method to estimate the behavior from the number of matching hints.

Khan *et al.*(2019) proposed an automatic method for estimating learning styles and affective states in an LMS. The learning styles are classified on FLSM dimensions as active-reflective, sensing–intuitive, visual–verbal, sequential – global. The confidence and independence dimensions of affective states are investigated. The patterns for identifying learning style and affective states were identified and then classified as high, low, or moderate existence. The total number of matching hints was computed based on information regarding a low or high occurrence to provide student preferences. The approach for the calculation of hints concerning learning style preferences and also the levels of respective affective states was based on a method developed by Graf and Kinshuk (2013) for detecting learning styles from behavioral patterns. The method applied the function:  $LS=(RT+RO)/2$  where,

- RT is the total time spent on a relevant learning object out of time spent on all objects.
- RO is the total number of relevant learning objects viewed out of all objects.

Based on this method, first, the time a learner is expected to take on a learning object is determined. The actual time taken to view a learning object relevant to an LS dimension is recorded. Then the respective ratios are computed. Similarly, the number of learning resources viewed is determined. The total number of learning objects relevant to an LS dimension is recorded. Finally, the average ratios are computed. Learning behavior is computed based on a 3-scale rule where values between 0 and 0.3 are weak, and 0.4 and 0.7 are moderate. 0.8 to 1.0 is strong. The estimation accuracy of



the method was evaluated using a study with 81 students. The results generated by the proposed method were compared with the learning styles and affective states questionnaires issued to the students. The study reported 0.81, 0.78, 0.76, and 0.74 precision on active–reflective, sensing intuitive, visual-verbal, and sequential, global learning styles. The study reported 0.82, 0.79 precision on confidence and independence dimensions of affective states. The results indicate high precision between 0.79 and 0.82 suggesting that the approach is appropriate for the identification of learning styles and affective states.

This thesis extends the work of Khan *et al.*(2019) by presenting a method for automatically estimating the student's learning styles and cognitive traits in an LMS from their interactions behavior with learning contents. The evaluation of the proposed method in this thesis was conducted using a study with 389 students who were taking an online course in an LMS.

### **2.6.3 The gap in the Previous Studies**

The previous studies reviewed in this chapter provide evidence that there is a lot of interest in learner behavior modeling. However, most of the studies considered have some limitations.

One critical observation made is that many of these studies, for example, Bernard *et al.* (2015) relied on manual tools such as the ILS questionnaire to evaluate the results of the automated learner modeling process. The problem with these instruments is psychometric flaws in measurement which are prone to errors and consume a lot of time to fill.

Although several personality models have been proposed, difficulties often arise when an attempt is made to select the appropriate technique.

There is evidence of overlap among the learning styles proposed. For example, Myers Briggs Type Indicator (Myers, 1995) Sensing (S) - Intuition (N) is also present in Felder-silver Model as

sensing – intuitive. Moreover, students have diverse ways of approaching studies and respond to instructions differently. This makes it difficult for a teacher to apply a specific learning style model.

Moreover, it is not possible for a teacher to routinely modify instruction methods to accommodate all learners with various psychological traits cited in the literature. These psychological theories are discussed with an emphasis on traditional face-to-face learning without considering computer-based classrooms.

Another problem with some learning models proposed is that they are based on a single student's behavior resulting in an incomplete or partial student modeling process and thus unreliable information.

In a systematic review, Namoun and Alshantiti (2020) pointed out that the state of predictive modeling of student behavior is not fully exploited and warrants further work. Similar sentiments were expressed by Umer et al., (2021) who called for the enhancement of the engagement dataset in learner modeling to cover data from more learning objects e.g. forums, and messages to identify more meaningful features that have some pedagogical influence in online learning. This implies learner modeling is an emerging body of research that requires more studies to be undertaken.

Analysis of recent literature reveals that Khan *et al.* (2019) presented a hybrid model for identifying relevant patterns of identification of students' affective states and learning styles in an LMS. The study investigated active–reflective, sensing–intuitive, sequential–global, and visual-verbal dimensions of FSLSM. The study also identified relevant patterns for confidence, independence, effort, and confusion dimensions of affective states, which led to the development of Affective States and Learning Styles Identification and Measurement' (AsLim). The tool is an example of a

hybrid model that identifies multiple learner behaviors in an LMS. Even though the study investigates a hybrid behavior, the evaluation of this approach was done by tracking student browsing behavior on a course hosted on Moodle LMS. Thus this method only tracked the learner behavior on a specific LMS platform and patterns cannot be generalized. In addition, there is a lack of a universal list of learner behavior indicators that can be used to profile and group learners in terms of learning style and cognitive behavior. While analyzing the results of a systematic literature review to discover the relationships between learning styles and learning objects, Assis et al.(2022) considered learning objects related to audio, animation and real-life applications to be related to sensing–verbal, reflective–visual, and global learning styles respectively. In contrast, Renato (2020) when developing an adaptive virtual system considered learning objects related to facts, forum, illustrations, and content navigation sequence to be related to sensing–verbal, reflective–visual, and global learning styles respectively. This confirms that there is little agreement on the navigation patterns that can be used as indicators of learning preferences.

This thesis addresses the gap by proposing a hybrid model which includes LS and CT variables to create a common vocabulary of concepts. This study proposes a method for designing and developing a hybrid learner behavior model integrating learning style and cognitive traits. The study intended to apply learning styles and cognitive traits which meet the following conditions:

- Have a significant impact on a student's learning process according to theories in education psychology
- Have pedagogical implications that can be applied in an LMS
- The descriptions can easily be inferred from records of student interactions with an LMS.

In the thesis, the modeling techniques suitable for learners with different learning styles and cognitive traits are defined with the corresponding mapping rules. The study provides an

automated hybrid learner behavior model integrating characteristics related to learning styles and cognitive trait models. The hybrid model is designed according to the descriptions of the dimensions that appear in the learning style and cognitive theories. The model comes with several benefits:

- It solves the problem of single student behavior resulting in an incomplete or partial student modeling process and thus unreliable information
- A hybrid model with more than one psychological theory solves the probe of the overlap among the learning styles
- The automated model solves the psychometric flaws that are associated with manual learner modeling methods.

## **2.7 Theoretical Framework**

The FSLSM, and cognitive behavior mainly CTM, is the main theoretical frameworks applied in this study. This is because human minds perform mental operations like computers. Information processing theory states that an individual registers sensory information taken from the environment, then processes and stores it. In education psychology, learning styles and cognitive traits are elements of the information processing theory.

FSLSM classifies learners as active-reflective, sensing-intuitive, visual-verbal, sequential - global. On the other hand, CTM focuses how people perceive, interpret, remember and think about what is learned. These internal mental processes are unique to each person. Working memory capacity, information processing speed, inductive reasoning ability, and association learning are some of the dimensions of CTM that are important for learning.

FSLSM is preferred since it accommodates the dimensions suggested in other theories. For instance sensing – intuition is also present in the MBTI (Myers, 1995). Active experimentation and reflective observation are also suggested in the Kolb learning model (Kolb, 2015). Visual, auditory, and read/write are also present in the VARK learning styles. This helps to create a universal model incorporating all major learning styles to ensure complete and reliable learner modeling information.

Furthermore, the LS dimensions such as FSLSM are easily inferred from LMS log records. For example, a student who views illustrations is assumed to have a visual learning style. A student who only views content with words is assumed to be having a verbal learning style.

This research further incorporated the main characteristics of CTM proposed in the literature such as information processing speed, working memory, inductive reasoning, and associative learning. Just like LS dimensions, only CT preferences that can be dealt with in LMS are included in the HYBRID model.

## **2.8 Conceptual Framework**

A conceptual framework defines the relevant variables for a study and maps out the relationships with each other.

### **2.8.1 Related Study**

A study conducted by Brusilovsky and Millán (2005) suggested users' backgrounds and preferences as variables that can be applied in an adaptive education system. Users' background defines previous experience with the subject. User preference is familiarity with the structure and the ability to navigate the system.

Magoulas and Dimakopoulos (2005) suggested cognitive or learning styles, user preferences, and interests. Cognitive or learning styles define a user's modes of perceiving and processing information. Interest defines the users' short-term goals. Preferences define users' likes and dislikes.

Rothrock *et al.* (2002) suggested user personality and cognitive style. The personality attributes can be introverted or extroverted. The authors recommended that it was appropriate to design an application for individuals with different characteristics.

Triantafillou, Georgiadou, and Economides (2009) reviewed different variables that can be used in adaptive educational systems. These variables can be classified as learner-dependent and learner-independent. Learner-dependent variables define a learner. They are associated with psychological theories that describe learner characteristics e.g. LS or CT. Learner-independent variables define the context of learner interaction with an LMS.

In the analysis and design of an adaptive education system, variables associated with learner interactions with an LMS are used as input for predicting learner behavior related to personality traits under investigation. Harrak, Bouchet, and Luengo (2019) carried out research to identify the relationship between students' questions and their profiles, using existing data and applied this analysis in predicting future student profiles in the next iterations of the course. The student profile based only on the questions asked were extracted from LMS records and used as input for a prediction model to create clusters based on attendance, student grade, number, and popularity of questions asked. It can be argued that in this study, the student profile extracted from an LMS represent the learner-independent variable. This is the predictor variable. The clusters created based on the learner behavior represent the learner-dependent variable. This is the outcome variable.

In another related study, Khan *et al.* (2019) introduced a rule-based method for extracting learning styles from an LMS as a mapping between the FLSM and learners' interactions with an LMS. The researcher examined visits to class exercises, time spent on exercises, and replies on the discussion forum to estimate LS. In Khan's study, the learner-independent variable group is based on navigation patterns matching particular descriptions of learning styles and affective states. The learner-dependent variables investigated in this case using FLSM dimensions were active-reflective, sensing-intuitive, sequential-global, and visual-verbal LS.

This thesis extends the previous works on identifying variables for adaptive education systems by highlighting the relationship between learner navigation records in an LMS and the psychological variables.

The study presents an integrated conceptual framework comprising variables related to LS and CT. It is envisaged that hybrid variables will lead to a better prediction outcome and more accurate learner personalization.

### **2.8.2 Definition of Conceptual Framework in the Current Study**

In the conceptual framework, this study suggests that learner interaction logs may be the best method for predicting LS and CT in an LMS. The following subsections describe the learner-independent and learner-dependent variables.

#### **Learner independent variables**

Independent variables in this study were derived from learner navigation patterns matching a particular dimension of FLSM and CTM.

These include the number of content pages visited, time spent on a relevant content page, the percentage amount of content covered, the number of attempts in online exercises, and marks

scored. The LMS used in the study observed log records of the learning object viewed by each student and the time spent. The LMS also recorded the number of quiz attempts and marks scored.

Independent variables related to LS were derived from:

- time spent on activities/number of visits to activities
- time spent on summaries/number of visits on summaries
- time spent on sub-topics/number of visits to sub-topics
- time spent on definitions/number of visits to definitions
- time spent on illustrations/number of visits to illustrations
- time spent on text/number of visits on text
- time spent on step-by-step navigation/number of step-by-step navigations
- time spent on introductions/number of visits on introductions

Independent variables related to CT were derived from:

- time spent revisiting/number of revisited pages
- time spent on examples/number of examples
- percentage learning content covered
- number of attempts at exercises
- marks scored on exercises

### **Learner dependent variables**

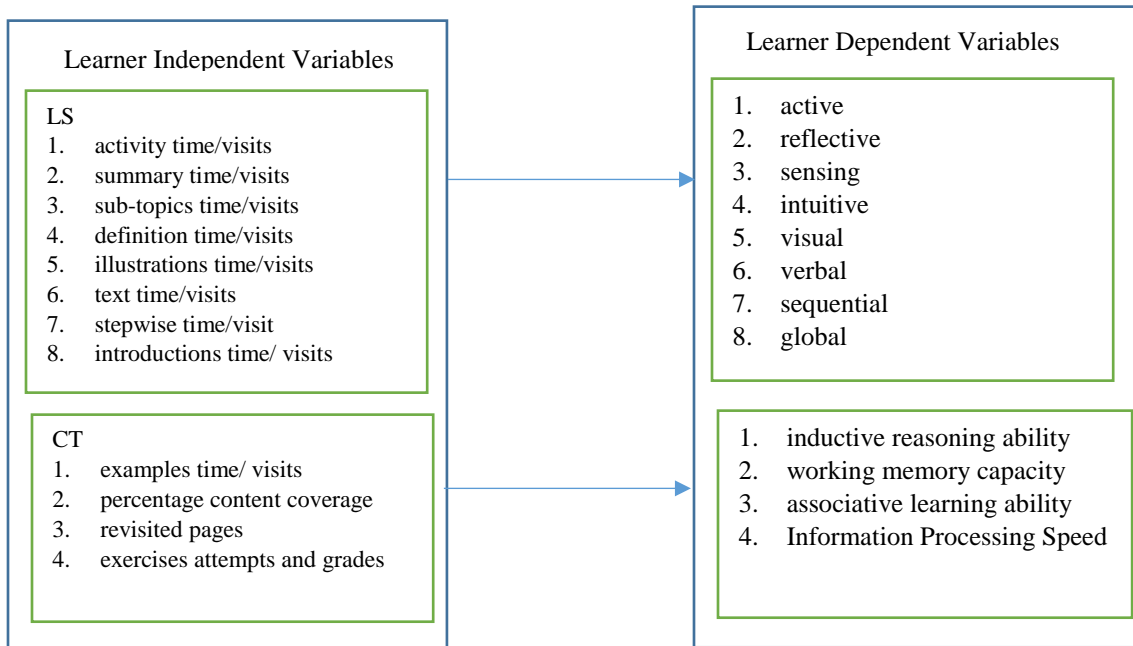
The dependent variables were derived from the personality models investigated in this study.

The dependent variables related to LS were derived from FLSLM which classifies a learner as active-reflective, sensing-intuitive, sequential-global, and visual-verbal.



The dependent variables related to CT were derived from CTM which classifies a learner based on working memory, inductive reasoning, associative learning, and information processing speed.

The cognitive trait can be classified as low, moderate, or high preference.



*Figure 2.1 Conceptual Framework*

Figure 2.1 illustrates the relationship between learner independent and dependent variables for both LS and CT.

The study makes the following assumptions concerning variables related to LS:

1. visits and time spent on contents with activities indicate active LS
2. visits and time spent on contents with summaries indicate reflective LS
3. visits and time spent on contents with sub-topics indicate sensing LS
4. visits and time spent on contents with definitions indicate intuitive LS
5. visits and time spent on content with illustrations indicate visual LS

6. visits and time spent on contents with text indicate verbal LS
7. visits and time spent on contents without skipping indicate sequential LS
8. visits and time spent on content with introductions indicate global LS

The study makes the following assumptions concerning variables related to CT.

1. visits and time spent on contents with examples indicate inductive reasoning ability
2. the percentage of content navigated forward out of the total indicates the working memory capacity
3. time spent revisiting pages indicates associative learning ability
4. attempts and grades on exercises indicate information processing speed

### 2.8.3 Operationalization of Variables

As shown in Tables 2.1 and 2.2, show records annotated to match the descriptions of FSLSM and CTM respectively.

Table 2.1. Operationalizing Variables related to FSLSM

FSLSM	Description of LS dimension	Learning objects considered	Pattern extracted	Classification criteria
Active	Tries out what is learned	Activity questions, assessment tests	Visits and time in minutes	active or reflective
Reflective	re-examines learned materials	summary, conclusion, revision	Visits and time in minutes	
Sensing	avoid challenging tasks but concentrate on details	Details e.g. topics, subtopics, explanations	Visits and time in minutes	sensing or intuitive
Intuitive	Prefer challenging tasks, but is uncomfortable with detailed contents	Descriptions	Visits and time in minutes	
Verbal	Reads text	Content with texts only	Visits and time in minutes	visual or verbal
visual	Likes viewing images	Images, illustrations	Visits and time in minutes	
Global	skips pages; interested in viewing the overall picture	introductory, content overviews, content outline	Navigation order, visits, and time in minutes	sequential or global
Sequential	Follows content step by step	introductions, content, summary pages	Navigation order, visits, and time in minutes	

Table 2.1 shows investigated patterns related to FSLSM. To measure LS, each LS dimension of FSLSM contains two opposite extremes, thus such as when patterns indicate the high end of a given dimension e.g. active; a low indicator shows an opposite pole of the same dimension e.g. reflective. The number of content pages viewed is counted as visits. The duration a page is viewed is time spent in minutes.

According to FSLSM, a student who likes trying out what is learned is an active learner; a student who takes time to meditate on the contents learned is a reflective learner. In this study, students who viewed more activity pages were considered to be active learners. Students who viewed more summaries, conclusions, and revision pages were considered reflective learners.

A sensing learner dislikes challenging tasks but is patient with details; an intuitive learner likes challenges but is impatient with details. A student who concentrates on content with details but avoids challenging tasks is a sensing learner. A student who likes tasks that are challenging is an active learner;

In this study, a student who viewed more topics and sub-topics with explanations and facts was considered a sensing learner. A student who viewed more content with descriptions and definitions was assumed to be an intuitive learner.

A student who views more learning materials with text is a verbal learner. A student who views more images is a visual learner. In this study, a student who viewed more pages with text was assumed to be a verbal learner; a student who viewed more content with images was assumed to be a visual learner.

A student who begins reading by looking at the whole picture of the content is a global learner; a student who follows the content step by step as presented is a sequential learner. In this study, a

student who viewed more pages containing content with introductions, overviews, and outlines was assumed to be a global learner; a student who skipped less content was assumed to be a sequential learner. Such a student would view introductions, then go to topics and sub-topics, then finish with summary pages and revision questions.

*Table 2.2. Operationalizing Variables related to CT*

<b>CTM dimension</b>	<b>Description of CT</b>	<b>Learning objects considered</b>	<b>Pattern extracted</b>	<b>classification</b>
Associative learning ability	Ability to relate new to existing knowledge learned earlier	Revisited pages, pages visited once	Visits and time in minutes	high, low, moderate
Working memory capacity	ability to concentrate on a reading task without losing focus	Order of navigating content. e.g. backward or forward	Visits and time in minutes	high, low, moderate
Inductive reasoning ability	understanding content and creating new ones from examples	Pages with content on examples	Visits, time in minutes	high, low, moderate
Information processing speed	How fast does a learner recall information read and make the correct decision	Exercises	Attempts, score	high, low, moderate

Table 2.2 shows the investigated patterns related to CTM. To measure CT, the current study introduced three measurement characteristics namely low, moderate, and high for each dimension.

This was done to indicate the cognitive strength of each characteristic.

The visit indicates the frequency of viewing a page with relevant learning content. Time indicates how long the page was viewed. Attempts indicate how many times within the exam schedule the answers are revised and submitted. A score indicates the grade a student gets on an online test.

According to CTM, an associative learner reads by creating a connection between new and previously acquired knowledge. In this study, a student who frequently viewed content pages that were already visited was assumed to be an associative learner. The frequency was rated as high, moderate, or low.

Working memory capacity checks to what extent a student can concentrate on a reading task without losing focus. In this study, a student who frequently visited content pages by navigating forward was assumed to be a sign of high working memory capacity. This shows the ability to concentrate on a reading task by keeping a limited amount of information read in the mind. A student who frequently viewed already visited pages was assumed to have a low working memory capacity.

Inductive reasoning checks to what extent a student can make concepts from examples. In this study, a student who frequently visited pages with contents on examples was assumed to have high inductive reasoning. Those who made fewer visits or none were assumed to have low inductive reasoning.

Information processing speed checks to what extent a learner can quickly retrieve the knowledge from memory and use the information to make the correct decision. In this study, a student who made a few attempts on an online test but scored a good grade was assumed to have a high information processing speed. A student who made multiple attempts with a score below average was considered to have a low working memory capacity.

From the operationalization of variables, a prototype estimating learning behavior was developed. Further, the dataset was analyzed with artificial neural networks and cluster algorithms to classify learners' in terms of LS and CT.

## **2.9 Summary**

This chapter has reviewed relevant literature on learner modeling methods.

In this chapter, it has been explained that a student model gathers information based on the system's beliefs about a student. Modeling is building and updating a system that gathers

information on student behavior when using the system. Most LMS only have data analytics components but lack intelligent features that can profile learner behavior such as LS and CT.

Information processing theory is the foundation of cognitive psychology that uses computers to represent the working mechanism of the human mind. This theory offers useful explanations of how the mind perceives and remembers information. Learning styles and cognitive trait models are the two individual behavior developed from information-processing theories.

Researchers have proposed automatic modeling methods where records using learner interactions are processed to infer their characteristics. To identify learner characteristics in an LMS, the patterns are extracted for analysis to create learner profiles.

Even though the studies presented in this chapter provide evidence that there is a lot of interest in learner behavior modeling, the majority are investigating either LS or CT but not both. Investigating a single behavior results in half-finished modeling. The proposed hybrid model is based on FLSM and CTM.

In the conceptual framework, this study suggests that learner interaction logs may be the best single method for predicting LS and CT in an LMS.

The current study proposes the use of an automated hybrid model comprising LS and CT. The study proposes a method for the analysis and interpretation of records of learner navigations in an LMS to identify LS and CT. The study identifies learning characteristics to be investigated and suggests information needed from students to identify learning preferences.

The next chapter discusses the research methodology applied in this study.

## **CHAPTER 3: RESEARCH METHODOLOGY**

In the previous chapter, the techniques reported in the literature for assessing learner behavior in an LMS are discussed. Since learning styles and cognitive trait models were not developed for computer-based learning, learner behavior has been traditionally measured using manual psychometric tools. The use of automated learner behavior modeling is done by monitoring the behavior and actions of learners during interactions with an LMS. This involves analyzing the records of learner interactions with LMS to infer the behavior. In all these studies, learners are classified by investigating the relationship between patterns extracted from LMS logs and psychological theories.

This chapter discusses the learner behavior modeling techniques used in this study. It describes the research design and strategy. It also outlines the process for extracting features from an LMS and matching them with the descriptions of the personality models investigated. Finally, it provides a discussion of the modeling process.

### **3.1 Research Design**

The purpose of this study was to design, develop and evaluate an automated model for predicting LS and CT from records of learners' interactions with an LMS. In this regard, the research design process was organized following a learner behavior modeling procedure suggested by Graf and Kinshuk (2008) and also reported by Khan *et al.* (2019). Figure 3.1 illustrates the research design process used in this study.

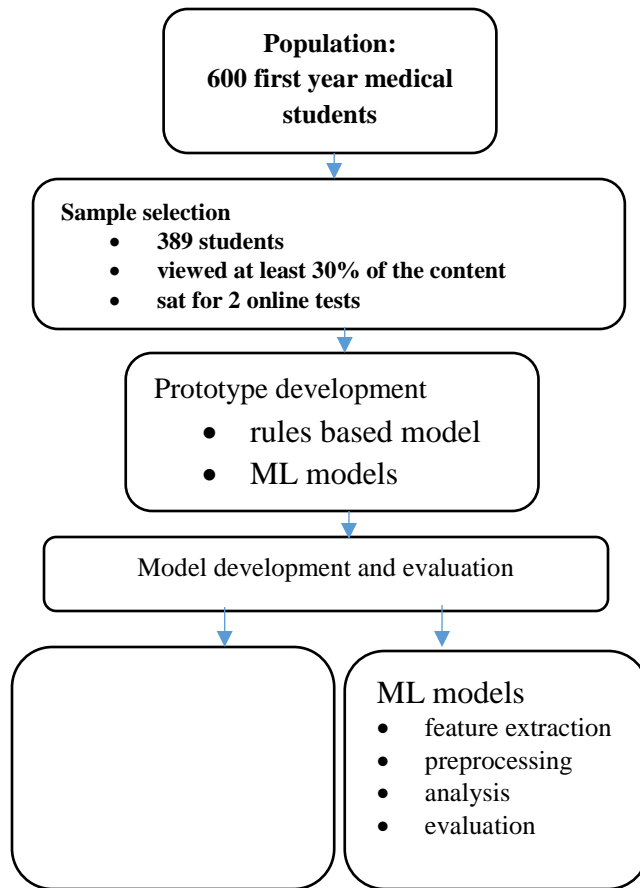
In both studies,

1. the study participants were selected from the student population who were accessing learning resources on an LMS
2. samples were selected from the participants based on a set of criteria e.g. minimal amount of learning content viewed, class participation, etc.
3. data were collected from records of learner interactions with an LMS
4. indicators of learning behavior were extracted from the records
5. the indicators were mapped to the psychological theories under investigation to estimate learning behavior
6. The estimation results were evaluated using case studies with students.

Thus the research design in the current study addressed the following questions:

- 1 How can feature sets for the automatic detection of learner behavior be extracted from LMS logs?
- 2 How can a computing model be used to detect learner behavior e.g. LS and CT in an LMS?
- 3 How can the above computing model be evaluated?





*Figure 3.1. Research design process*

### **3.1.1 The Population of the Study**

The target population was 600 first-year students at the University of Nairobi for the academic year 2017/2018 and 2018/2019 taking a preclinical medical course. The students were registered for a common course unit in behavioral science taught by the department of medical psychiatry. The course was divided into ten lecture modules namely psychology, social processes, foundations of human behavior, social processes, sociology, and social psychology, anthropology, physical illness causing the behavior, neurosciences, and healthcare systems.

The lecture modules were written by the course lecturers and converted to the shareable content object reference model (SCORM) format. This is a content standard for packing and reusing

learning materials in different LMS platforms (Varlamis and Apostolakis, 2006). The learning contents were hosted on Claroline open-source LMS. Each student was assigned a username and password. The study focused on these groups because the students were actively using the LMS to access learning resources. The students accessed the learning materials for a 15-week semester. The lecture modules were read online through self-study with minimal student-lecturer interactions. The students sat for two multiple-choice tests online. The questions were set from the contents covered in the sociology and psychology modules. The LMS tracking tools recorded the access statistics for each student.

### **3.1.2 Sample Selection**

The purposive sampling technique was used to select a smaller group of participants who met the threshold of the study. This is following Chauncey (2014) who indicated that when using this sampling technique, the individuals who meet the specific goals of the study are sought out. To select the sample from this population, browsing records of the students on the lecture modules hosted on the LMS were inspected. The content access tracking feature of the LMS recorded the statistics of the learning objects visited by each student. Participants were included in the study if the LMS records indicated they had accessed at least 30% of the content and sat for two online tests. Participants were excluded from the study if the LMS records indicated they had accessed less than 30% of the contents irrespective of whether they sat for continuous assessment tests. Using these selection criteria, a sample of 389 students were selected to represent the entire population. The students selected for the study were contacted through the course lecturers.

### **3.1.3 Data Collection**

The content navigation records for each student were extracted from the LMS log. The content navigation patterns indicated the type of content viewed by each student. The content objects were identified as introductions, definitions, sub-topics, activities, illustrations, grades on tests, attempts on tests, and percentage content access. The indicators of FSLSM and CTM were identified from the navigation patterns. Using this extracted data; a rule-based model, an artificial neural network model, and a cluster model were designed and implemented. This addressed the second research question on how feature sets for the automatic detection of learner behavior could be extracted from LMS logs.

### **3.1.4 Model Development**

The development of a rule-based model was done using evolutionary prototyping. The method was preferred because of the unclear user requirements in the study. The algorithms for estimating LS and CT were designed and developed using PHP/ MYQL scripting language. The models were implemented to predict LS and CT. The rule-based model was further enhanced to filter and present learning materials relevant to LS and CT identified for each student. The development steps for the rule-based model are later discussed in sections 3.3 and 3.4 of this thesis.

To compare the performance of the model against machine learning methods, further analysis was done using artificial neural networks and cluster algorithms.

The development of an artificial neural network model for predicting LS and CT was done using R software version 3.5.1. The dataset labels were prepared to match the dimensions of FSLSM and CTM. The content labels were assigned as the area before content, definitions, content, area after content, activities, visual content, assessment score, attempts, and average percentage content

coverage. The dataset with input values gathered from the sociology and output from psychology modules were entered into the network model. First, cluster models with variables related to LS and CT were each developed separately. Then a hybrid model with variables related to both LS and CT was developed. The development of an artificial neural network model is discussed in section 3.6 of this thesis.

The development of a cluster-based model was done using R software version 3.5.1. The dataset labels were prepared to match the dimensions of FLSM and CTM. First, cluster models with variables related to LS and CT were created and trained separately. Thereafter, a hybrid cluster model with variables related to both LS and CT was created and trained together. The model development process addressed the third research question on how a computing model could be used to detect learner behavior e.g. LS and CT in an LMS. The development of the cluster model is discussed in section 3.7 of this thesis.

### **3.1.5 Model Evaluation**

The model evaluation process addressed the fourth research question on how to measure the performance of the model created.

#### ***Evaluating rule-based model for identifying LS and CT***

Data was collected from the students to evaluate the prediction accuracy of the rule-based model. The students were briefed on the objective of the study. A total of 200 students answered 44 questions from the index of learning styles (ILS) questionnaire, and thereafter they attempted online psychometric tests. The tests were adapted from Cambridge Brain Sciences (*Cambridge Brain Sciences*, 1998). This is an online website used by psychologists to administer cognitive

tests. After students completed the ILS questionnaires and the online psychometric tests, their results were scored.

The responses from the ILS questionnaires were computed using a questionnaire scoresheet to identify LS. The mark sheets for psychometric tests were downloaded from the psychometric website. The data were then entered into a Microsoft Excel spreadsheet for analysis using statistical software.

The results were prepared in discrete categories: low, moderate, and high. Interrater reliability analysis tests were done using SPSS software version 23 to evaluate the agreement between the results predicted by the model and the psychometric methods. Model evaluation steps are discussed in section 3.5 of this thesis.

### ***Evaluating machine learning models for predicting LS and CT***

The network training was done repeatedly by modifying the hidden layers. The output generated at each iteration was compared against the expected results to measure the prediction accuracy of the network. The machine learning metrics precision, recall, f-score, and area under the curve were used to measure the performance of the network. Section 3.6 of this thesis discusses the neural network model development and evaluation process.

The performance of the cluster model developed was evaluated using the silhouette coefficient to measure the compactness, connectedness, and separation of the partitions created. Section 3.7 of this thesis discusses the cluster model development and evaluation process.

### **3.1.6 Validity and Reliability of Model Evaluation Instruments**

#### ***The index of Learning Styles (ILS)***

The index of Learning Styles (ILS) questionnaire was used to collect data for evaluating the results related to LS. ILS is the main measurement tool used to assess LS based on FSLSM. The questionnaire has 44 questions designed by Felder and Silverman at North Carolina State University, USA to find FSLSM preferences. Learners complete the questionnaire by putting a circle "a" or "b" to choose one answer only for each question. One must answer all 44 questions. If both "a" and "b" seem to apply to a respondent, then the one that applies more frequently is chosen.

The web-based version of the ILS has been used in several published studies, some of which include data reflecting on the reliability and validity of the instrument. For example, Felder and Spurlin (2005) reviewed published studies on the reliability and validity of the ILS instrument and reported that test-retest correlation coefficients for all four scales of the ILS instrument were significant at the 0.05 level or better. Cronbach alpha coefficient values for each dimension of the ILS instrument were remarkably consistent with one another. These studies confirm that the ILS is suitable for assessing learning styles.

#### ***Online Cognitive Tests***

The Cambridge Brain Sciences (*Cambridge Brain Sciences*, 1998) tests were used to evaluate the results related to CT. The cognitive tests have been evaluated in psychology studies and proven to be efficient measures of baseline cognitive capacity. For example, in a study conducted by Levine *et al.*, (2017), the results gathered via this method were comparable to those of a standard paper and pencil neuropsychological method. The cognitive test outperformed the standard test of cognitive abilities in a study conducted to investigate the mental capacity of elderly patients

(Brenkel *et al.*, 2017). In another related study conducted by Hampshire *et al.* (2012), the performance of Cambridge Brain Sciences was highly predictive of reasoning and problem-solving abilities as indexed by classic tests such as the Cattell Culture Fair test. This implies Cambridge's brain of sciences tests is suitable for assessing cognitive behavior.

### **3.1.7 Presentation of the Evaluation Results**

#### **Presenting Evaluation Results for Rule-Based Model**

The evaluation results for the rule-based model were presented using tables and graphs generated by SPSS software for data analysis.

#### ***Frequency Distribution Table***

A frequency table is a statistical analysis method for listing items under investigation considering the number of times the item is observed. In the current study, frequency tables were used to show the distribution of LS and CT by strengths. For example Table 4.1 shows the frequency distribution for LS as provided in the next chapter. Tables 4.2, 4.3, 4.4, and 4.5 show the frequency distribution for CT.

#### ***Cross-Tables***

Cross-tabulation is a statistical method for analyzing the correlation between variables. Cross-tables were used to illustrate the results predicted by the rule-based model against the psychometric methods. The method was applied in the current study to analyze and understand the relationship between the results predicted by the model and the manual psychometric tools. The cross tables were used to present both the results related to LS and CT. Tables 4.6, 4.7, and 4.8 show cross-tabulation results for LS. Tables 4.9, 4.10, 4.11, and 4.12 show cross-tabulation results for CT.

### ***Graph***

A pie chart was used to illustrate the distribution of participants who took part in the evaluation by gender. Figure 4.1 show the distribution of participants by gender illustrated using a pie chart.

### **Presenting Evaluation Results for the ANN Model**

The evaluation results for the ANN model were presented using tables and graphs. The features for generating tables and graphs are easily available as implementation functions in R software for neural network model development.

### ***Confusion Matrix Table***

The confusion matrix table was used to assess the ratio of correctly predicted observations to the total predicted observations for the results related to LS and CT. Figure 4.2 shows the confusion matrix table for the results related to LS and CT variables.

### ***Artificial Neural Network Plot***

The plotted artificial neural models indicating different input nodes and hidden layers for variables related to LS, CT, and hybrid were generated and presented. The results for the area under the curve were illustrated using a graph. Figures 3.8.1, 3.8.2, 3.8.3, and 3.8.4 show the plotted neural network models.

### **Presenting evaluation results for cluster Model**

The evaluation results for the cluster-based model were presented using tables and graphs.

The features for generating tables and graphs are easily available as implementation functions in R software for cluster model development.



### ***Silhouette tables***

The silhouette analysis results for variables related to LS, CT, and hybrid (all variables) were summarized in tables. Each table indicated the analysis results for the four algorithms used k-means, PAM, Diana, and Hierarchal algorithms. The cluster placements for students indicating the size and average silhouette were generated by the model and presented in tables. Tables 4.18, 4.19, and 4.20 show the cluster tabulated results for variables related to LS, CT and a hybrid of both respectively.

### ***Silhouette plot***

The plot is a visual illustration to show the closeness of the points in a given cluster to the points in the neighboring clusters. The cluster silhouette plot for indicating average silhouette widths for variables related to LS, CT, and hybrid (all variables) was generated and displayed. Figures 4.4, 4.5, and 4.6 show silhouette plot results for variables related to CT, LS and a hybrid of both respectively.

## **3.2 Learning Contents Hosted on LMS**

The contents of the learning modules used in this study are divided into main chapters called units. Each unit shows the outline of the sections to be covered and the objectives to help the student get a global view of the chapter. Each unit is divided into sections. The sections show objectives, the definition of concepts, and the outline of the topics to be covered to help the students understand the scope. The definition of concepts states the technical terms used in the text and their meanings. Each section ends with a summary of the topics discussed and self-assessment questions to enable a student to revise what is learned. Each section is further divided into topics and sub-topics to be covered. Topics and sub-topics are concepts and facts discussed in detail. The learning materials are made up of both text and visual content. Each topic has activity questions to enable learners to

test their understanding of the concepts. Appendix 3 shows a section of an electronic learning module hosted on Claroline LMS. The electronic learning content page displayed on an LMS has a navigation window to the left displaying a list of topics and sub-topics as shown in appendix 4. The content is displayed on the right when a page is selected. The heading and sub-headings for the content pages have numbers indicating the level.

### **3.2.1 Patterns Extraction from contents**

To extract patterns related to learner interactions with the contents, the tables in the Claroline LMS database with relevant records of learner interactions with the contents were first identified. In the Claroline LMS database, the default tables are assigned names using the prefix `cl_`. The database has approximately 40 default tables. In addition, each course unit created in the LMS is assigned a set of independent tables for managing learning resources. Such tables are assigned the prefix identifying the course code. The behavioral science course used in the current study was assigned code HCH100 in the LMS. The LMS automatically assigns the tables associated with the course unit the prefix HCH100\_. In the LMS database, `cl_user` is a default table for storing user records in the Claroline database. The `cl_log` is a default table for tracking and storing the transaction records in the system. The `HCH100_lp_module` is the learning path table for storing the SCORM learning module for the HCH100 course. `HCH100_lp_rel_learnPath_module` is the table that stores the sections of the module and relates them to the student records in the `cl_user` table. `HCH100_lp_user_module_progress` is the table that captures and stores the records of user interactions with the learning module. The table shows the sections accessed, time spent, and date for each student. `HCH100_qwz_tracking` table stores mark scored and attempt on tests. Attempts are records indicating how many times a student revised and resubmitted answers after the first

submission. These tables provided data for learner interactions with the contents which were used in this study.

SQL search queries were written to extract patterns from the learning path content tracking records in the LMS database. The keyword related to the sections, sub-topics, and topics of the learning modules were used to identify features related to LS and CT theories under investigation in the study. The queries applied keywords identified from the pages of the learning content hosted in the LMS as features for identifying the LS and CT. The queries searched the tracking tables in the LMS database and returned the access records for each student. The search queries are shown in appendix 36.

- A search query with a keyword containing the term ‘activity/activities’ was used to identify patterns related to active LS. The query returned the number of relevant pages viewed and the total time spent in minutes by each student who visited pages with activity questions. The search query for reflective LS was based on keywords containing the terms ‘summary’, ‘self-assessments’, and ‘revision’. The queries were used to generate the records on the number of relevant pages viewed and the total time spent in minutes by each student.
- A search query with a keyword identifying topics and sub-topics accessed by a student was used to identify patterns related to sensing LS. The query returned the number of relevant pages viewed and the total time spent in minutes by each student. These pages were searched using the heading number formats as keywords to return the number of pages viewed and the time spent by each student. The query for extracting patterns related to intuitive LS was based on phrases containing the keyword ‘definition’. The query returned the number of relevant pages viewed and the total time spent in minutes by each student.

- A search query containing keywords ‘figure’, ‘illustration’, ‘picture’, etc. was used to identify patterns related to visual LS. The query returned the number of relevant pages viewed and the total time spent in minutes by each student. The records that did not appear in the search records were considered patterns for verbal LS.
- A search query containing keywords unit, section, module, etc. was used to identify patterns related to global LS. The query returned the number of relevant pages viewed and the total time spent in minutes by each student. The query for identifying patterns related to sequential LS was based on the step-by-step access to introduction sections, followed by topics/sub-topics then finishing with summary and end-of-unit tests.
- A search query returning the number of the same pages accessed on different times and dates by each student was used to identify patterns related to associative learning ability. The study assumed that displaying different access times and dates for each page indicated a student revisited such content.
- A search query containing the keywords ‘example’ or ‘case studies’ was used to identify patterns related to inductive reasoning ability. The query returned the number of pages visited and the time spent by each student.
- A search query indicating single access dates and times in ascending order against a student considering percentage content coverage was used to identify patterns related to working memory capacity. The study assumed that displaying a single time and date for each page indicated a student visited such content once. Such a student concentrated on a learning task.
- A search query returning the number of attempts and marks scored on exercises was used to identify patterns related to information processing speed.

### 3.2.2 Patterns Extracted from an LMS Log Records

The navigation records were extracted and analyzed to identify patterns related to LS and CT. Table 3.1 shows sample log data extracted from an LMS. The numbers in the first column represent individual learners.

- Introductions indicate the frequency of visiting the course, unit, and section overviews.
- Definitions indicate the frequency of visiting content with new meanings.
- Topics and sub-topics indicate the frequency of visiting content with details and facts.
- Summaries indicate the frequency of visiting content with summaries and conclusions.
- Activities indicate the frequency of visiting content with learning activities.
- Illustrations indicate the frequency of visiting content with images.
- Time on introductions indicates the duration in minutes a student spent viewing content with overviews.
- Time on definitions indicates the duration in minutes a student spent viewing contents with new meanings and terms.
- Time on topics and subtopics indicates the duration in minutes a student spent viewing contents with details and facts.
- Time on summaries indicates the duration in minutes a student spent viewing contents with learning summaries and conclusions.
- Time on activities indicates the duration in minutes a student spent viewing content with learning activities.
- Grade on tests shows the average marks scored in all online tests graded out of 100.
- Attempt on tests shows how many times and student revised and resubmitted answers. The minimum attempt is 1 and the maximum is 3.

- Percentage content coverage shows the amount of content a student accessed continuously without interruptions.

Table 3.1. Sample Log Data Extracted from an LMS

learner	Introduction	Definitions	Topics and	Summaries	Activities	Illustrations	Time on illustrations	Time on introductions	Time on definitions	Time on topics	Time on summaries	Time on activities	Grade on tests	Attempts on test	Percentage content coverage
1	86	60	20	80	0	50	71	80	297	560	156	0	65	1	59
2	86	100	40	60	0	100	74	27	86	231	14	0	79	1	57
3	71	100	80	100	100	100	36	11	48	144	15	36	66	2	75
4	100	100	100	80	100	100	44	16	51	150	16	0	81	1	61

### 3.3 Learner Behavior Model

This section describes the design, development, and evaluation of the learner behavior model. This study used the name CogniLearn to represent the model. The prototype is based on FSLSM and CTM. The model can identify learning behavior and display relevant learning content matching the learning styles and cognitive traits of an individual student.

To meet the needs of this research, the CogniLearn model was extended with the following features as shown in figure 3.2.

- Pattern extraction component with Learning Style Generator (LSG) and Cognitive Trait Generator (CTG)
- Storage repositories for keeping patterns and learner preference
- Modeling component for estimating LS and CT preferences
- Content matching component for generating relevant learning resources that match LS and CT preferences
- Display interface for viewing learning preferences and the contents

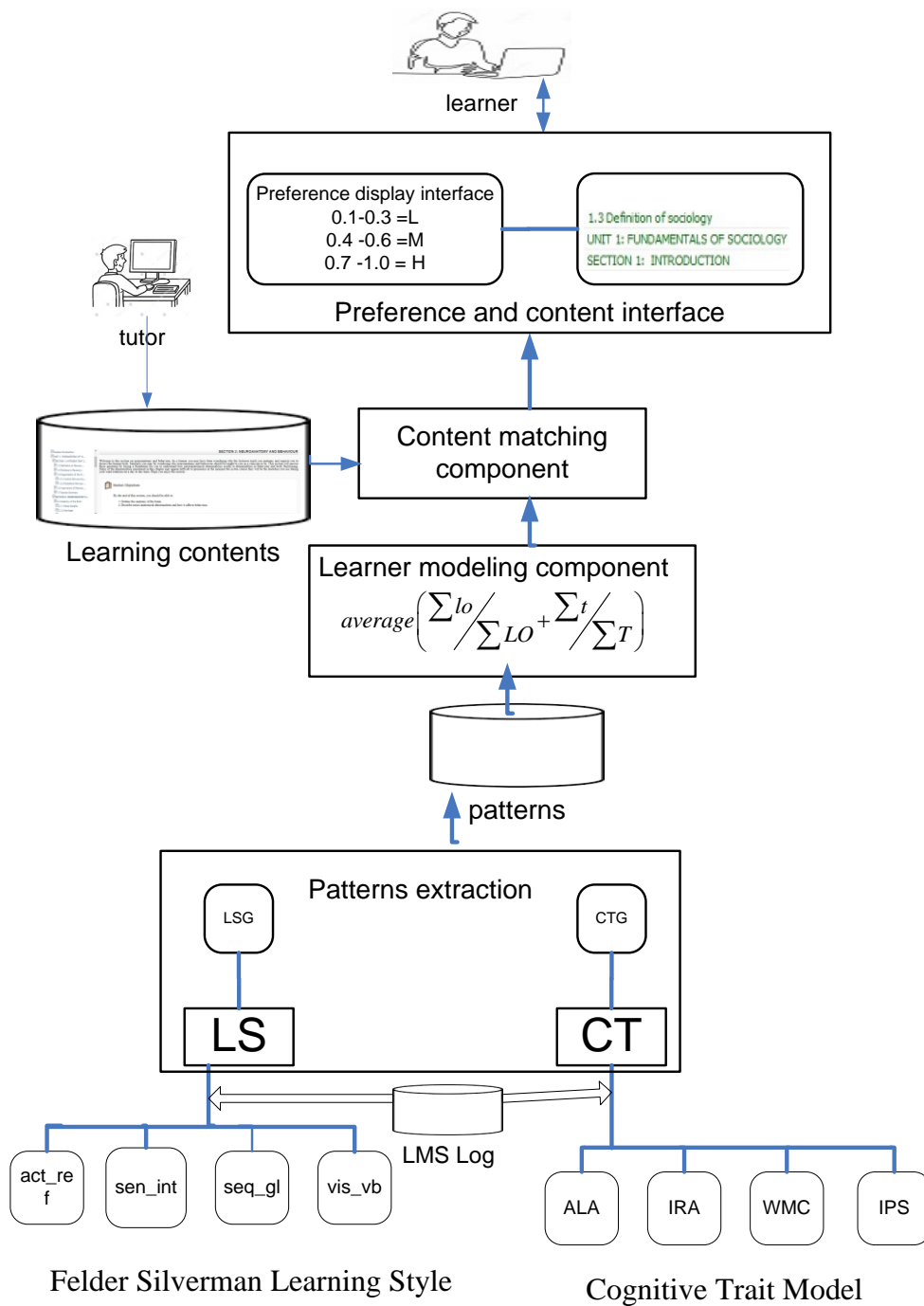


Figure 3.2. CogniLearn Model Architecture

Once the patterns of learners' interactions with an LMS are extracted from the log file, the records are matched with the relevant descriptions of LS and CT models. From this, the LS and CT for

each student are estimated and displayed via the model interface. The relevant patterns associated with LS and CT are fetched from the LMS log by the extraction engines. The LSG and CTG engines compute the LS and CT from the patterns extracted. The LS and CT for each student are estimated and mapped to 3 item scale: 0.1 - 0.3 low, 0.4-0.6 Moderate, 0.7 – 1.0 high, and 0.0 – no preference. The model matches learning contents following the individual characteristics estimated. Results are displayed via the model interface. The following discussions consider the main components of the model.

### **3.3.1 LMS Log**

Data used in this study was collected from Claroline LMS which is distributed under the Gnu Public License (GPL) and is built on PHP/MySQL. The LMS is developed on a web environment, to be available free on the Internet. With the LMS, instructors can set courses and enroll students. The system has an exercise module for administering online tests. The content management modules are used to host learning resources. The system generates statistics indicating the learning content accessed and time spent. Claroline LMS is easy to use having a user-friendly interface for instructors and students. When students access learning resources, records of interactions with the contents are captured and stored in the database. Sample data from LMS are presented in Appendix 8.

### **3.3.2 Pattern extraction component**

The LMS captures access records when students are browsing the content. The relevant LS and CT records are extracted from the LMS database by a pattern extraction engine. The patterns are content objects accessed, how long a learner stays on a learning object, navigation order, marks, and attempts on the exercises. The patterns related to LS are shown in Table 2.1. The patterns related to CT are shown in Table 2.2. Learning Style Generator (LSG) extracts patterns related to



LS. Cognitive Trait Generator (CTG) extracts patterns related to CT. In this study, records of learner interactions with the LMS were extracted and stored in a separate database. From these, patterns matching LS and CT were extracted.

The patterns and preference storage components were implemented as a relational database. Appendix 11 shows the ERD of the database. The database was implemented using the MySQL management tool. The content storage section is used for keeping learning materials posted by an instructor. The preference storage section keeps the LS and CT for each student computed by the estimation component.

### **3.3.3 LS and CT Estimation Components**

An estimation component fetches relevant patterns from the storage to compute LS and CT for a student based on the access patterns. As shown in Tables 3.2 and 3.3, the access patterns are activities in an LMS.

The computation method was adapted from the methodology for calculating the learner's profile considering the mapping provided by Graf, Kinshuk, and Liu (2009), and repeated by (Khan et al., 2019) to develop an automatic student model for identifying learning styles and affective states in web-based LMS. The actions of learners on an LMS are used to get hints about their preferences before using a rule-based method to estimate learning styles and affective states.

In this study, the matching preferences for FSLSM such as active-reflective, sensing-intuitive, sequential-global, and visual-verbal were considered. The computation for each pair of the LS preference is based on the following steps:

- i. Total time spent viewing learning objects relevant to an LS dimension (t) (e.g. active or reflective) out of total time spent on all objects related to the pair under investigation (T)

$$\frac{\sum t}{\sum T} \text{ (e.g. active + reflective)}$$

- ii. The frequency of visits on a learning object relevant to LS dimension (lo) (e.g. active or reflective) out of total objects related to the pair under investigation accessed (LO)  $\frac{\sum lo}{\sum LO}$

(e.g. active + reflective)

- iii. The average is computed from the results in steps 1 and 2.  $average\left(\frac{\sum lo}{\sum LO} + \frac{\sum t}{\sum T}\right)$

- iv. Map the results to the item scale: 0.1 - 0.3 Low, 0.4 - 0.6 Moderate, and 0.7 – 1.0 High

Table 3.2. LS Estimation functions

FSLSM	Relevant Learning Object	Pattern extracted	Estimation function for average ratio.
<b>Active (a)</b>	Activities	object visits (o) and time (t)	$average\left(\frac{ta}{ta+tr}, \frac{oa}{oa+or}\right)$
<b>Reflective(r)</b>	summaries, conclusion, revisions	object visits (o) and time (t)	$average\left(\frac{tr}{ta+tr}, \frac{or}{oa+or}\right)$
<b>Sensing (s)</b>	Topics and subtopics topics	object visits (o) and time (t)	$average\left(\frac{ts}{ts+ti}, \frac{os}{os+oi}\right)$
<b>Intuitive(i)</b>	Definitions, meanings	object visits (o) and time (t)	$average\left(\frac{ti}{ts+ti}, \frac{oi}{os+oi}\right)$
<b>Verbal(v)</b>	Textual content	object visits (o) and time (t)	$average\left(\frac{tv}{tv+tv_i}, \frac{ov}{ov+ov_i}\right)$
<b>Visual(vi)</b>	Images, illustrations	object visits (o) and time (t)	$average\left(\frac{tvi}{tv + tvi}, \frac{ovi}{ov + ovi}\right)$
<b>Global(g)</b>	Introductions, overviews, outline	object visits (o) and time (t)	$average\left(\frac{tg}{tg + tsq}, \frac{og}{og + osq}\right)$
<b>Sequential(sq)</b>	Introductions, topics, summary	object visits (o) and time (t)	$average\left(\frac{tsq}{tg + tsq}, \frac{osq}{og + osq}\right)$

The matching pairs for each CT dimension were considered. For example, pages that were revisited against those visited once were considered for *associative learning*. For *working memory capacity*, forward navigation against backward was considered. The following steps were used to estimate the CT:

i. For each CT dimension, time spent on relevant content objects (t) (e.g. revisited vs visited once for associative learning or forward vs backward navigation) out of total time spent on all objects (T)  $\frac{\sum t}{\sum T}$  (e.g. revisited + visited once for associative learning or forward + backward navigation)

ii. For each CT dimension, the frequency of visits on a relevant object (lo) (e.g. revisited vs visited once for associative learning or forward vs backward navigation) dimension out of total objects accessed (LO)  $\frac{\sum lo}{\sum LO}$  e.g. revisited + visited once for associative learning or forward + backward navigation)

iii. The average is computed from results in steps 1 and 2  $average\left(\frac{\sum lo}{\sum LO} + \frac{\sum t}{\sum T}\right)$

iv. The results are mapped to 3 item scale: 0.1 -0.3 Low, 0.4-0.6 Moderate, and 0.7 – 1.0 High to get estimated preference

The summation  $average\left(\frac{ti}{T} + \frac{o}{O}\right)$ . Time spent on relevant objects (ti) and total objects accessed (o) out of the total time (T) and total objects (O) were considered for inductive reasoning ability.

The summation  $average\left(\frac{a}{A}, \frac{s}{S}\right)$  attempts (a), and score (s) out of the total attempts (T), and the total score (T) for relevant objects were considered for information processing speed.

Table 3.3 shows the estimation procedure for each CT dimension.

Table 3.3. CT Estimation Functions

CTM dimension	Relevant Learning Object	Pattern extracted	Estimation function
Associative learning	Revisited pages, pages visited once	Visits and time	$average \left( \frac{trp}{trp+ts}, \frac{orp}{orp+os} \right)$
Working memory capacity	Forward navigation, reverse navigation	Visits and time	$average \left( \frac{tr}{tr+tf}, \frac{or}{or+of} \right)$
Inductive reasoning ability	Examples, revision, and exercises	Visits, time	$average \left( \frac{ti}{T}, \frac{o}{O} \right)$
Information processing speed	Exercises	Attempts, score	$average \left( \frac{a}{A}, \frac{s}{S} \right)$

### 3.3.4 Content Matching Component

Figure 3.3 illustrates the content matching component which provides learning content for each student relevant to the estimated LS and CT estimates.

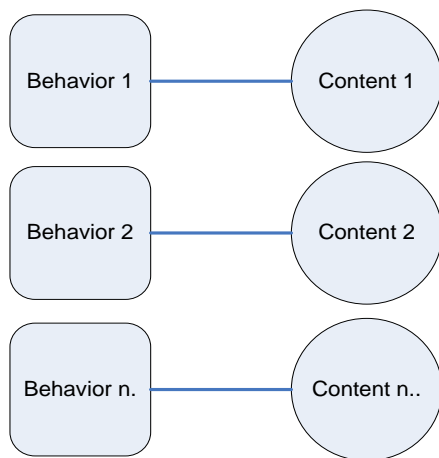


Figure 3.3. Matching content with relevant LS and CT.

The system uses profile-based filtering, to estimate the LS and CT preferences of each student recommending appropriate learning content. In CogniLearn, the estimated LS and CT become part of the student's learning profile. Through content-based filtering, the system uses learning object

metadata to match desired contents with LS and CT. Content-based recommendation systems match the description of the learning content and the profile of the user.

Learning materials can be dynamically filtered from a knowledge base to customized learning materials for each student centered on their knowledge and preferred mode of learning method (Sridharan *et al.*, 2021). As soon as CogniLearn receives a request for content recommendation, the learning object database is searched and relevant learning materials matching LS and CT are selected. Table 3.4 describes the link between learning preferences and learning content. The learning content type matches the respective LS and CT.

*Table 3.4. Matching the Learning Preferences with Appropriate Contents*

<b>Content-type</b>	<b>Matching LS and CT</b>
Learning activity	Active
Revision and Summary	Reflective
Explanations/facts	Sensing
Definitions	Intuitive
Overview, outlines	Global
Lessons	Sequential
Illustration/image	Visual
Text/audio	Verbal
External resources	Associative learner
Reading comprehension	High Working memory capacity
Short notes/key points	Low working memory capacity
Example/case study	High inductive reasoning ability

Active learner prefers learning activities; reflective views revision and summary pages. Sensing learner prefers explanations and facts; intuitive like definitions of concepts and terms. A global learner prefers course overviews and outlines; a sequential learner prefers lessons organized in a systematic order. A visual learner prefers illustrations while a verbal prefers text and audio-based content. Associative learner prefers external resources. A learner with a high working memory capacity can concentrate on a reading task. A learner with high inductive reasoning prefers

examples and case studies. While uploading learning materials, a course tutor has a provision for specifying content types. The content-type feature enables the model to match the learning preferences of each student with appropriate content. Once a student access the CogniLearn, the LS and CT preferences are recorded in the learner profile, which then is compared to the learning content type. The content and profile-based filtering feature then compare the information from the learning contents and the student's profile before recommending the appropriate content.

### **3.3.5 Display interface**

The display interface is the component for visualizing LS and CT. It is the visualization of a learning map that presents LS and CT preferences. The interface can display the distribution of preferences for each learner. The hybrid behavior of the model allows the combination of both LS and CT to be displayed for each learner. Learners can therefore search, identify their learning characteristics and make comparisons with classmates. The display interface also enables a student to see contents matching individual LS and CT. The model interface functionality is presented in Appendix 14.

## **3.4 Model Prototyping**

This research used the evolutionary prototyping method to develop a learner behavior model. Using evolutionary prototyping as a development method, a sample is first constructed and then verified by the stakeholders. Subsequent prototypes are developed with additional functionalities until the final product is ready (Anjum *et al.*, 2019; Ariffin *et al.*, 2020).

Evolutionary prototyping was preferred in this study to help address the anticipated complexity of a diverse range of LS and CT models on how to define and incorporate different and conflicting psychological variables into acceptable requirements. The evolutionary prototyping used in the model development adapted the procedure used by Agustiono (2018) to develop A Spatial Web-

Based System For Land Management. Developers in evolutionary prototyping implement the most essential working model of the system and then continue its refinement based on user feedback until the requirements are met. Agustiono investigated the development of a Land Management Information System using the evolutionary prototyping method. First, requirements elicitation and analysis were done by reviewing related regulations and laws relating to soil and land protection. Second, the prototype building activities such as architecture, database, module, and user interface were designed according to user requirements. These activities were verified and modified until a working prototype was made.

In carrying out the model prototyping, first, the requirement analysis was conducted on existing relevant literature regarding the learning style and cognitive theories. This involved investigations on features and patterns related to the varying learning styles and cognitive traits. The studies outlining the relationship between the learning objects and learning theories were reviewed to construct the mapping between content navigation patterns and psychological models.

Thereafter, a web-based prototype that enabled students to visualize and measure their LS and CT was developed. The model displays learning materials matching LS and CT for each student. Finally, the modeling accuracy of the prototype was evaluated using a case study with students. Appendix 9 shows the system flow diagram. Appendix 10 presents the evolutionary prototyping steps.

### **3.4.1 Requirement Specifications**

In the requirement analysis phase, the researcher reviewed related studies on the LS and CT identification in an LMS to get the system specifications. The previous studies give direction on how different types of learning contents are related to psychological theories (Khan *et al.*, 2019; Assis *et al.*, 2022). For example, audio-learning contents fit verbal learning styles. Therefore, it

can be affirmed that a student who views audio content is well-fitted to a verbal learning style. Learning objects with illustrations fit the visual learning style. Therefore, it can be affirmed that a student who views much of the illustrations is well-fitted to a visual learning style. The prototype captures records of learner interactions with an LMS. The records are kept in a storage database. The modeling component would extract the records from the storage component to estimate LS and CT. The adaptation component would match learning contents with relevant LS and CT. The output of the computations is displayed by the interface component of the model. The requirements were presented in research seminars and conferences to get feedback. Based on the feedback, the specifications were continuously modified to improve the performance of the model. The browsing records and the patterns that matched the LS and CT used in this study were identified. The model was expected to compute the preferences of learners from the browsing patterns identified. The storage component maintains the database of extracted patterns. The processing phase estimates the LS and CT. The output phase provides the interface where a student enters a registration number to know the LS and CT. The model also displays relevant content matching the LS and CT for each student.

### **3.4.2 Prototype Development**

Once the requirements were gathered, the researcher started to develop a prototype. This involved designing the model architecture, database, estimation engine, content-matching component, and user interface based on the requirements. These activities were repeated until a working prototype was ready. At each development stage, the prototype was continuously modified to improve the input, processing, storage, and output.

The MYSQL database was used to design and implement the storage component of the model. The PHP scripting language was used to implement the estimation component of the model. The



two implementation tools are open-source software available for free. Using the evolutionary prototyping method, the model was continuously built and tested.

### ***Implementing input***

The user input interface was implemented using PHP scripting language and HTML forms. The web-based model has a login interface that allows a user to enter a username and password for access. There are three user levels namely an administrator, a lecturer, and a student.

The system has a text field where learning contents can be captured. The system also has a provision for file upload. For each learning object, the system has a drop-down list for selecting a learning content type. The course lecturer can type or copy and paste contents on a text field. The course lecturers can also upload a document file. A student is provided with content and preference display links.

### ***Implementing storage***

The model storage was designed using a relational database method. Appendix 11 shows Entity-Relationship Diagram (ERD) for the Cognilearn Model Database. The database was implemented using the MYSQL tool. All the storage tables were created and the logical relations were tested. At each stage, tests were done to confirm if it was possible to insert, delete, update, edit, and sort records. The functionalities were continuously improved. Appendix 12 shows the data dictionary. The data dictionary describes each table included in the entity-relationship diagram and the fields.

### ***Process***

As shown in Table 3.2, for LS the number of learning objects  $n$  relevant to a particular LS dimension e.g. active – reflective, sensing – intuitive, sequential – global, visual-verbal, and time spent were considered. Figure 3.4.1 shows a sequence of logical steps for estimating LS.

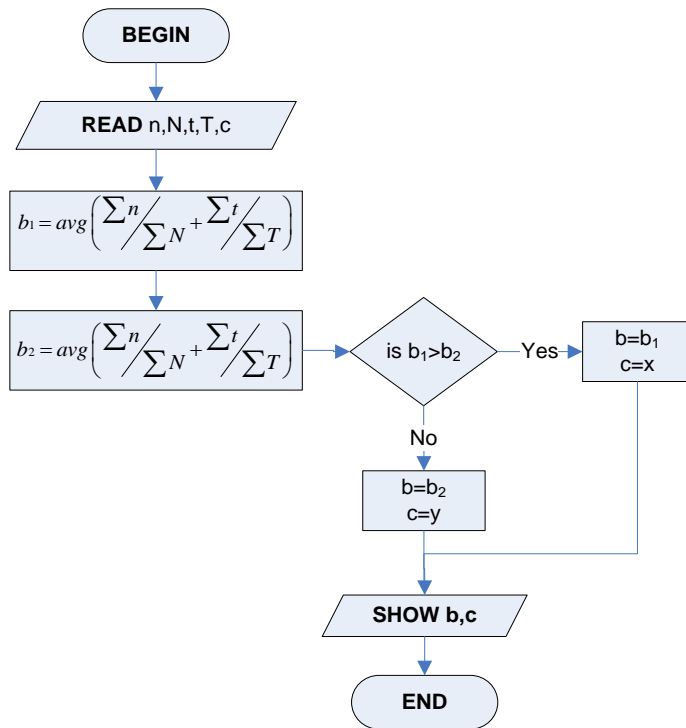


Figure 3.4.1 LS Estimation Algorithm

In figure 3.4.1:

- n stands for relevant visited pages that match a given LS dimension e.g. active - reflective.
- N stands for the total number of content pages visited.
- t stands for the time spent on pages that match a given LS dimension.
- T for the total time spent on all content pages.
- c stands for the relevant learning content matching a particular LS.
- b stands for a learning behavior e.g. active or reflective LS.
- x and y are values indicating specific content types matching estimated CT.

The algorithm for LS detection and content allocation works as follows:

For a pair of LS dimensions e.g. active – reflective, sensing–intuitive, visual-verbal, sequential–global,

BEGIN

Step 1: take variables n, t, c, b

Step 2: assign values to each variable

Step 3: for each LS dimension e.g. active – reflective, sensing – intuitive, visual-verbal, sequential – global, calculate the average  $(n,t)$  for each dimension in the pair and store the values in  $b_1$  and  $b_2$  respectively. Take note that  $b_1 - b_2$  represent matching pairs e.g. active – reflective.

Step 4: if  $b_1 > b_2$ , store the value of  $b_1$  in  $b$  and the content matching it in  $c$ , or else store the value of  $b_2$  in  $b$  and the content matching it in  $c$ . The system assumes there is no preference if  $b_1=b_2$ . This is a balanced preference and is not used to identify a student (Felder and Soloman, 2017)

Step 5: display the values of  $b$  and  $c$ .

END

As shown in Table 3.3, for ALA, the revisited contents  $n$ , and time spent  $t$  out of the total visits  $N$  and time  $T$  on all contents were considered. The number of pages visited in the forward direction  $n$  and time spent  $t$  out of the total visits  $N$  and time  $T$  on all contents was considered for WMC. The number of contents with examples viewed  $n$ , and time spent  $t$  out of the total visits  $N$  and time  $T$  on all contents was considered for IRA. The number of attempts  $a$  and grades  $s$  on all exercises was considered for IPS. Figure 3.4.2. below shows a sequence of logical steps for estimating CT.

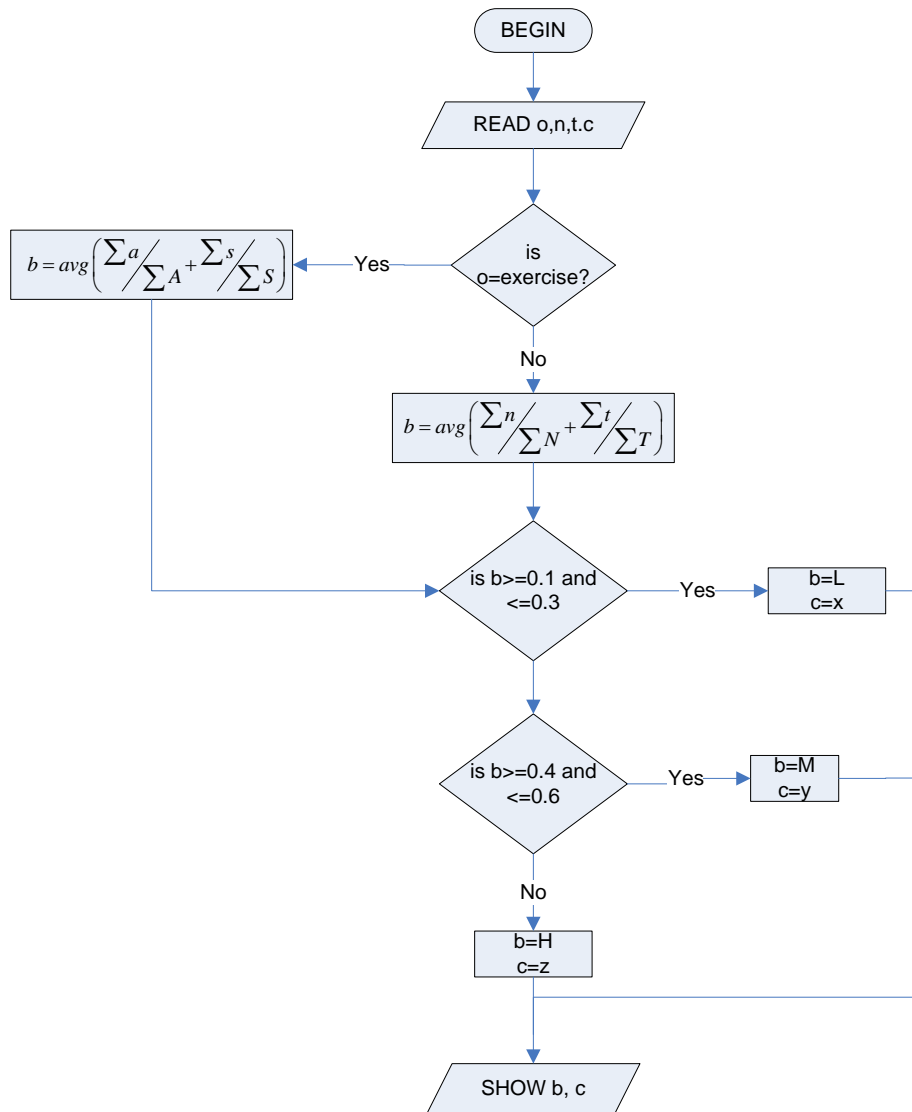


Figure 3.4.2. CT Estimation Algorithm

As shown in figure 3.4.2. above, the letter

- a stands for the number of attempts made on exercises
- s stands for grade scored on exercises
- n stands for relevant visited pages that match a given CT e.g. ALA.
- N stands for the total number of content pages visited.
- t stands for the time spent on pages that match a given CT.
- T stands for the total time spent on all content pages.
- c stands for the relevant learning content matching a particular CT.
- b stands for a learning behavior e.g. high, moderate, or low ALA.
- o stands for a learning object with relevant content e.g. exercise page, examples, etc.
- x, y, and z are values indicating specific content types matching estimated CT

The algorithm for CT detection and content allocation works as follows:

For a set of CTs e.g. ALA, IRA, IPS, WMC

BEGIN

Step 1: take variables o, n, t, a, s, c, b

Step 2: assign values to each variable

Step 3: for each CT e.g. ALA, IRA, IPS, WMC, if the value stored in o is exercise, calculate the average (a,s) and store the values in b. If not, calculate the average (n,t). An object exercise is a learning object used to get the values for IPS.

Step 4: Store the values for step 3 in b

Step 5: For each value of b,

if b is between 0.1 and 0.3, store the content x in c,

if b is between 0.4 and 0.6, store the content y in c,

if b is between 0.7 and 1.0, store the content z in c.

The letters x, y, and z represent the learning contents relevant to a student with low, moderate, or high cognitive traits.

Step 5: display the values stored in b and c.

END

The prototype implementation and testing were done using PHP/MYSQL scripting language. The system was inspected at each stage to ensure it could select LS and CT records from the storage, perform the correct computations and display appropriate contents to each student. The results

were checked to ensure they were accurate. The numerical results between 0.0 and 1.0 were displayed. For matching learning styles such as active–reflective, sensing–intuitive, sequential–global, and visual-verbal, the results for each pair would add up to 1.0. Furthermore, learning content matching each behavior was also displayed. The computation process was checked and the module was repeatedly refined. The system flowchart for the CogniLearn model appears in appendix 9. The program code segments for the CogniLearn model appear in appendix 13. The model interface and functionalities are illustrated in appendix 14.

### **3.5 Model Evaluation**

The evaluation of the model was done to confirm its prediction accuracy. This was carried out by administering manual psychometric tools to students in a classroom environment. These results were then compared against those generated by the model.

A sample of 200 students participated in the survey. The students were taking a degree course in bachelor of medicine and bachelor of surgery. The students had satisfactorily accessed the learning materials online and attempted all online tests.

To evaluate the LS results generated by the prototype, Index of Learning Styles Questionnaires (ILS) was administered to students. This is a standard measurement tool for the identification of LS based on FSLSM. The students filled out the questionnaires and returned them. The questionnaire was analyzed to identify the LS for each student.

To evaluate CT results generated by the prototype, online cognitive tests were administered to students on the LMS. The tests were adapted from Cambridge Brain Sciences (*Cambridge Brain Sciences*, 1998). This is an online website used by psychologists to administer cognitive tests.

### **3.5.1 Index of Learning Style Questionnaire**

The index of Learning Styles (ILS) questionnaire is the main measurement tool used to assess LS based on FSLSM. The questionnaire has 44 questions designed by Felder and Silverman at North Carolina State University, USA to find FSLSM preferences. Learners complete the questionnaire by putting a circle "a" or "b" to choose one answer only for each question. One must answer all 44 questions. If both "a" and "b" seem to apply to a respondent, then the one that applies more frequently is chosen. Appendix 15 shows the ILS questionnaire.

The scoring is done by placing a "1" in the appropriate spaces and then adding up the columns to get the total. The smaller total is subtracted from the larger one, writing the difference and the letter (a or b) with the larger total. For example: If the total is 3 for a, and 8 for b; therefore, the difference is  $8 - 3 = 5$ , with b being larger total so you would enter 5b. Results of 1-3 are mild, 3-5 moderate, and 9-11 strong preference for either dimension. Mild means well balanced; moderate indicates learning in an environment favoring such; strong infer to having difficulty learning in an environment not supporting the LS.

### **3.5.2 Online Cognitive Tests**

The Cambridge Brain Sciences provide tests for measuring cognitive behavior such as short-term memory, reasoning, attention, and verbal ability. The cognitive tasks provided have been evaluated by psychological studies e.g. investigating the relationship between sleep and intellectual abilities (Fang *et al.*, 2021), the influence of background noise in the office on concentration, and verbal reasoning performance (Liu, He and Qin, 2021).

While many websites are administering online cognitive tests, this study created MCQ tests similar to those administered by Cambridge Brain Sciences on an LMS to collect data on associative learning ability, inductive reasoning, and working memory capacity. The mental speed test used

to collect data on information processing ability was adapted from the psychology testing website (Perina *et al.*, 2021). Similar tests were created on an LMS and administered to evaluate the results related to CT.

The study administered the four categories of cognitive tests on a university LMS. Paired associate tests, spatial span tests, abstract reasoning tests adapted, and mental speed tests were used to measure the cognitive behavior of the students.

Paired associate tests were used to test the associative learning ability of the students. The paired-associate tests tested the ability to remember and recall specific events, paired with the context in which they occurred. The MCQ questions administered required participants to remember what had previously been seen, along with the location. For each MCQ, a graphical image with a set of boxes appeared on the screen opening, one after another, revealing the objects inside. The number of objects and boxes increased as the student progressed with the questions. The students were required to remember which object appeared in a given box by selecting the correct answer from the available choices.

Spatial span tests were used to test the working memory capacity of the students. The spatial span tests measured the ability to remember the relationships between objects in space. The MCQ questions administered required the participants to remember the correct sequence in the boxes that appeared over a period, by selecting the correct answer from the choices provided. For each MCQ question, a graphical image with boxes flashing in sequence appears on the screen and then disappears. The number of objects and boxes increased as the student progressed with the questions. Performance was indicated by the number of correct answers.



Abstract reasoning tests were administered to test the inductive reasoning ability of the students. The abstract reasoning tests measured the cognitive ability to apply rules to information to arrive at a logical conclusion. The MCQ tests required reasoning about the features of several shapes to draw conclusions on which shape did not conform with the rest. For each MCQ, a graphical image with shapes was provided. The students were required to select the odd one out by choosing the correct answer from the choices provided. The number of objects and boxes increased as the student progressed with the questions. Performance was indicated by the average number of correct answers.

Mental speed tests were used to test the information processing speed. The questions were adapted from the psychology today website (Perina *et al.*, 2021). The mental speed tests measured how quickly a student could process information and make decisions based on the available information. For each timed TF question a graphical word/image pair was provided. A student was expected to follow the specific instructions to attempt the questions. A student was expected to click the True button if the pair matched or False in case the pair is not matching. The answer would be reversed if the word "opposite" appeared. The number of objects and boxes increased as the student progressed with the questions. Performance was indicated by the number of correct answers. Sample cognitive assessment tests are illustrated in appendix 18.

### **3.5.3 Data Collection and Analysis Procedure**

The participants were briefed on the research aim to make them comfortable. The participants were guided on how to fill out the questionnaires and answer cognitive tests on an LMS.

To collect data on LS, the Index of Learning Styles (ILS) questionnaire was the main measurement tool used to assess the LS of the respondents based on FSLSM. A total of 200 ILS questionnaires were administered to students who participated. Each respondent was asked to answer the 44

questions in the questionnaire. The participants were instructed to follow the guidelines provided by the ILS (Felder and Soloman, 2017) and select the most preferred LS. The learning styles were computed as described in the FLSM ILS questionnaire.

To collect data on CT, 4 online cognitive tests were created on an LMS. Each cognitive test determines a specific cognitive trait. Each set of tests was a multiple choice type (MCQ) scheduled to last for one hour. The LMS automatically graded the answers out of 15 when a student submitted them upon finishing. Paired associate tests were used to evaluate associative learning ability. Spatial span tests were used to evaluate the working memory capacity. Abstract reasoning tests were used to evaluate inductive reasoning ability. Mental speed tests were used to evaluate information processing speed.

Before administering the online tests, the students were briefed on the procedure. Each student then sat on a computer, accessed the LMS, and attempted the 4 tests under the close supervision of the research team. The results were extracted and coded on an excel spreadsheet. A student who scored below 7 out of 15 was considered as having a weak cognitive trait. A student who scored between 7 and 10 was considered as having a moderate cognitive trait. A student who scored above 10 was considered as having a strong cognitive trait. The data for LS and CT were coded on an excel spreadsheet and analyzed further with SPSS software (version 23). The analysis was based on interrater reliability analysis to compare the results generated by the model against those gathered by psychometric methods.

### 3.5.4 Interrater Reliability Analysis

In learner behavior modeling, there is a need to frequently determine how results generated by the model are in agreement with those generated through psychometric tools such as learning style questionnaires. Some recent studies have used different methods to evaluate the accuracy between learner behavior prediction models and psychometric tools. Some of the previously used methods are correlation analysis (Ozdemir *et al.*, 2018; Park, 2017; Dvorak and Jia, 2016; Ghadirian, Fauzi Mohd Ayub and Salehi, 2018), precision and recall (Liyanage, Gunawardena and Hirakawa, 2014; Sweta and Lal, 2016). In this study, data collected during the evaluation was prepared in discrete categories such as: mild or low, moderate, and high. Since such data required specific statistical methods to assess reliability, the kappa statistic was selected as the data analysis method.

Kappa statistic is the metric method for assessing the agreement between two raters (Cohen, 1960). It indicates the proportion of agreement beyond that expected by chance, that is, the achieved beyond-chance agreement as a proportion of the possible beyond-chance agreement. It takes the

form:  $k = \frac{P_o - P_c}{1 - P_c}$  where  $P_o$  is the proportion of observed agreements and  $P_c$  is the proportion of agreements expected by chance.

Using kappa statistics, the interrater reliability test was done to measure the agreement between the behavior gathered through the Index of Learning Styles Questionnaires (ILS), Online Cognitive Multiple Choice Questions, and those estimated by the model.

Table 3.5. Interrater Reliability Test

Prediction model	Psychometric methods		Total
	True	False	
True	A	B	$g_1$
False	C	D	$g_2$
Total	$f_1$	$f_2$	N

Two methods are used to predict the LS and CT of a student As shown in table 3.5. The critical point here is how well these ratings agree. Using a 2 x 2 contingency table as illustrated by Sim and Wright (2005), cells a and d indicate the number of cases in which the two methods agree. Cells b and c indicate the number of cases in which the two methods of prediction disagree.

For this study, the corresponding marginal totals for the predictions using the psychometric method are  $f_1 + f_2$ . The corresponding marginal totals for the predictions using the model are  $g_1 + g_2$ .

The proportion of observed agreement  $p_0 = \frac{(a + d)}{n}$

The proportion of expected agreement  $p_c = \frac{\left(\frac{f_1 \times g_1}{n}\right) + \left(\frac{f_2 \times g_2}{n}\right)}{n}$

Substituting into the formula:  $k = \frac{p_0 - p_c}{1 - p_c}$

According to Landis and Koch (1977), the kappa values range from -1 to 1. The values below 0 show that there is no agreement. The values indicate the level of agreements as contained in each bracket 0.0 - 0.20 (slight), 0.21 - 0.40 (fair), 0.41 - 0.60 (moderate), 0.61 - 0.80 (substantial), and 0.81 - 1.00 (perfect agreement). This test was conducted to determine how well the implementation of the model was in agreement with the traditional methods. To determine how the results generated by the model were in agreement with the psychological methods, the reliability analysis to determine consistency among raters was done using the Kappa statistic method. SPSS software (version 23) was used for the analysis. Section 4.1 discusses the results of the model evaluation.

### 3.6 Learner Modeling using Artificial Neural Network Analysis

To compare the closeness of the results predicted by CogniLearn against a supervised machine learning method, the current study employed the ANN and cluster algorithms to infer the relationship between LS, CT, and the behavior of the students in an LMS. ANN requires little statistical training, can detect complex nonlinear relationships between dependent and independent variables, and has multiple training algorithms (Tu, 1996).

Artificial Neural Network (ANN) is characterized by a set of interconnected nodes (Rojas, 1996). The nodes received inputs, then process them to generate output. The information flow between nodes is facilitated by interconnections. As shown in figure 3.5 below, the incoming signals are converted into a single outgoing and then broadcast to other neurons. The activation of an ANN neuron is determined when the inputs are multiplied by weights and then computed by a mathematical function.

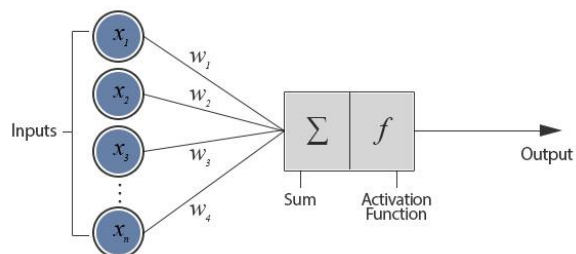


Figure 3.5. Neural Network Architecture

The training was carried out on the network using the following steps:

1. Present the network with training data set with input and output labels
2. Compare predicted against desired output
3. Modify the weight so that network produces better output.

The records gathered from an LMS were entered into the system as input to predict LS based on data gathered from ILS. In this study, the dataset extracted from the LMS log was used to train the ANN. The ANN was trained to predict the LS and CT of each student based on FSLSM and CTM. This study investigated only the WMC and IPS dimensions of CTM.

The records of learner interactions with contents on psychology were used as input for the network. The records of learner interactions with contents on sociology were used as output for the network. This is illustrated in Table 3.7. The dataset was further processed using if-else rules to get the output values. First, the patterns of learner interactions were extracted from LMS logs. The patterns further went through a pre-processing stage and cleaning to remove missing values. The training was done using R-studio language. The input dataset was entered into the network and trained to classify learners according to LS and CT. Accuracy, precision, recall, f-score, and area under the curve (AUC) were used to evaluate the prediction ability of the model. The network training steps are further described below.

### **3.6.1 Data Encoding**

The log data were processed further for ANN training. Table 3.6 shows sections of learning modules that were labeled with codes to facilitate ANN training.

- ABC stands for ‘area before contents’. These are pages that introduce the learning materials.
- Pages with the definition of concepts are labeled definitions (D).
- C stands for ‘contents’. These are pages with facts and concepts discussed in detail.
- A stands for ‘activities’. These are pages with activity questions.

- AAC stands for ‘area after contents’. These are pages with summaries and revision questions based on the contents discussed.
- V stands for ‘visual’. These are pages with contents showing illustrations.
- ASCORE stands for ‘assessment score’.
- Attempt shows how many times within the allocated time a student submitted and revised answers.
- AVG stands for the ‘overall percentage of content covered. These are the percentage of the total learning module pages viewed.

Table 3.6. Sample Dataset for Log data extracted from an LMS

Learner	Introduction	Definitions	Topics and	Summaries	Activities	Illustrations	Time on illustrations	Time on introductions	Time on definitions	time on topics	Time on summaries	Time on activities	Grade on tests	Attempts on test	Average percentage content coverage
	ABC	D	C	AAC	A	V	VT	ABCT	DT	CT	AACT	AT	ASCORE	ATTEMPTS	AVG
1	86	60	20	80	0	50	71	80	297	560	156	0	65	1	59
2	86	100	40	60	0	100	74	27	86	231	14	0	79	1	57
3	71	100	80	100	100	100	36	11	48	144	15	36	66	2	75
4	100	100	100	80	100	100	44	16	51	150	16	0	81	1	61

The dataset was further labeled to match the dimensions of FSLSM. The labels are a-r,s-i,v-r,g-s where:

- a is active,
- r is reflective,
- s is sensing,
- i is intuitive,
- s is sequential,
- g is global,
- v is visual,
- r is verbal.

This is an alternative to using the ILS questionnaire which students must fill out and return. This method addressed the challenges related to the administration of manual tools.

From Table 3.6, the variables related to active LS are activities and exercises; those related to reflective LS are summaries, conclusions, and revisions. The variables related to sensing LS are topics and sub-topics; those related to intuitive LS are definitions and meanings. The variables related to verbal LS are textual contents; those related to visual LS are images and illustrations. The variables related to sequential LS are based on the content navigation order; those related to global LS are introductions, overviews, and outlines. With regards to CT, the variable related to working memory capacity (WMC) is AVG which stands for the average content coverage. The variables related to information processing speed (IPS) are a score i.e. marks scored and attempts. The LS is inferred using if-else rules as shown:

IF A, AT > AAC, AACT then a ELSE r.  
IF C, CT > D, DT then s ELSE i.  
IF VS, = 100% then v ELSE r (for reader or verbal).  
IF ABC, ABC\_T, AAC, > ABC,C,AAC then g ELSE s.

CTM dimensions were encoded using if-else rules as low, moderate, and high. The letter L stands for low, M is moderate, and H is high.

IPS:  
IF Attempts>1,L,IF Attempts=1 and Ascore<69, M, else H.  
IPS  
WMC:  
IF AVG<50,L,IF AVG>=50 and AVG<60,M, else H.  
WMC.

IPS stands for information processing speed and WMC is working memory capacity.

The rules generated 67 possible combinations of LS and CT.



Figure 3.6 illustrates LS represented by the first four attributes. The last two attributes indicate CT. In this example, the student shows a preference for active, intuitive, global, and reflective LS. In terms of CT, the student has high IPS and low WMC.

student	a/r	s/i	s/g	v/r	i	w
1	a	i	g	r	H	L
2	r	s	s	v	M	L
3	a	i	g	r	H	M

Figure 3.6. Example of LS and CT Output Generated

Input data for training the ANN was generated from the records of the student's interactions with the psychology module. See Table 3.7. The records of students' interactions with the sociology module were used to infer the training output for the ANN.

Table 3.7. Input and output variables

	INPUT	LABEL	Description of the content object
Input dataset from psychology module	X <sub>1</sub> :	ABC	area before content
	X <sub>2</sub> :	D	Definitions
	X <sub>3</sub> :	C	Content
	X <sub>4</sub> :	AAC	area after content
	X <sub>5</sub> :	A	Activities
	X <sub>6</sub> :	V_S	visual content
	X <sub>7</sub> :	ABC_T	time spent on the area before content
	X <sub>8</sub> :	D_T	time spent on definitions
	X <sub>9</sub> :	C_T	time spent on content
	X <sub>10</sub> :	AAC_T	time spent on the area after content
	X <sub>11</sub> :	A_T	time spent on activities
	X <sub>12</sub> :	Ascore	assessment score
	X <sub>13</sub> :	Attempts	number of assessment attempts
	X <sub>14</sub> :	AVG	average % content access
	OUTPUT		
Output dataset generated using if-else rules from the sociology module	X <sub>15</sub> :	LS	Learner Behavior: FLSM and CTM dimensions LS{a: active, r: reflective, s: sensing, i: intuitive, s: sequential, g: global, v: visual, r: verbal, L: low, M: moderate, H: high}}

For ease of training with ANN, combinations of LS and CT were given a unique number ranging between 1 and 67. In the example given below, the first student has preferences for reflective, sensing, sequential, and visual learning LS.

For CT, the student has moderate WMC and IPS. Such LS and CT combinations are represented by a numerical value of 3.

Student	LB	code
1	rssvMM	3
2	asgvMH	5
3	assvHH	8
4	rssvMH	9
5	rssvML	7
6	asgvLH	12

Table 3.8 shows a sample dataset with 17 rows of records. Each row indicates an individual learner.

The dataset has 14 predicting and one response variable LB which stands for learning behavior.

*Table 3.8. Sample Dataset*

LEARNER	ABC	D	C	AAC	A	V_S	ABC_T	D_T	C_T	AAC_T	A_T	Score	Attempt	AVG	LB
1	86	100	40	60	0	100	21	26	100	9	0	68	1	84	1
2	100	100	60	80	100	100	29	331	320	18	0	68	1	82	1
3	71	100	60	60	100	100	15	48	239	25	15	62	1	85	1
4	57	100	80	80	0	100	20	111	1178	33	0	67	1	74	1
5	29	20	20	20	0	0	23	41	431	0	0	70	1	73	2
6	100	100	100	80	100	50	12	30	76	12	8	73	1	84	2
7	86	100	100	100	0	50	28	60	193	19	23	76	1	79	2
8	86	60	20	80	0	50	80	297	560	156	0	65	1	59	3
9	86	100	40	60	0	100	27	86	231	14	0	79	1	57	4
10	71	100	80	100	100	100	11	48	144	15	36	66	1	75	5
11	86	80	60	80	0	100	6	219	775	18	0	58	1	0	6
12	43	40	20	40	0	100	4	15	73	13	0	52.5	2	21	6

Table 3.9 shows the frequency distribution for learner behavior categories generated from the analysis. The minimum population in the distribution of 1 % has 3 students. The maximum

population in the distribution of 9.6 % has 30 students. The highest number of students is seen in category 18 with 9.6%. Category 7 has 8 %, followed by 21 at 6.4 %, and 25 at 4.8%.

Table 3.9. Frequency Distribution for Learner Behavior

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	3	1.0	1.0	1.0
	3	4	1.3	1.3	2.3
	5	7	2.3	2.3	4.5
	7	25	8.0	8.0	12.5
	8	4	1.3	1.3	13.8
	9	3	1.0	1.0	14.8
	12	5	1.6	1.6	16.4
	13	7	2.3	2.3	18.6
	15	5	1.6	1.6	20.3
	17	16	5.1	5.1	25.4
	18	30	9.6	9.6	35.0
	19	18	5.8	5.8	40.8
	21	20	6.4	6.4	47.3
	22	7	2.3	2.3	49.5
	25	15	4.8	4.8	54.3
	26	3	1.0	1.0	55.3
	27	3	1.0	1.0	56.3
	29	8	2.6	2.6	58.8
	30	3	1.0	1.0	59.8
	31	5	1.6	1.6	61.4
	33	5	1.6	1.6	63.0
	34	2	.6	.6	63.7
	35	4	1.3	1.3	65.0
	36	2	.6	.6	65.6
	38	12	3.9	3.9	69.5
	40	2	.6	.6	70.1
	41	3	1.0	1.0	71.1
	42	3	1.0	1.0	72.0
	48	9	2.9	2.9	74.9
	52	5	1.6	1.6	76.5
	53	2	.6	.6	77.2
	54	3	1.0	1.0	78.1
	55	11	3.5	3.5	81.7
56	11	3.5	3.5	85.2	
58	5	1.6	1.6	86.8	
59	2	.6	.6	87.5	
60	4	1.3	1.3	88.7	
61	10	3.2	3.2	92.0	
62	8	2.6	2.6	94.5	
63	5	1.6	1.6	96.1	
67	9	2.9	2.9	99.0	
69	3	1.0	1.0	100.0	

### 3.6.2 Data Pre-processing

The dataset was cleaned by removing the missing values and outliers. Data normalization was done using the feature scaling method. The numeric data was converted to scale [0,1] using the data scaling function shown below.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

In the function above;

- $x'$  is the normalized dataset
- $x$  is the original
- $x_{\min}$  is the minimum
- $x_{\max}$  is the maximum

The scale function of the r programming language was used to normalize the dataset. The sample code is shown below:

```
max_data <- apply (data [, 1:14], 2, max).
```

```
min_data <- apply (data [, 1:14], 2, min).
```

```
data_scaled <- scale (data [, 1:14], center = min_data, scale = max_data - min_data).
```

Further analysis was done to establish how the response variable LB was distributed. Figure 3.7 shows the results obtained.

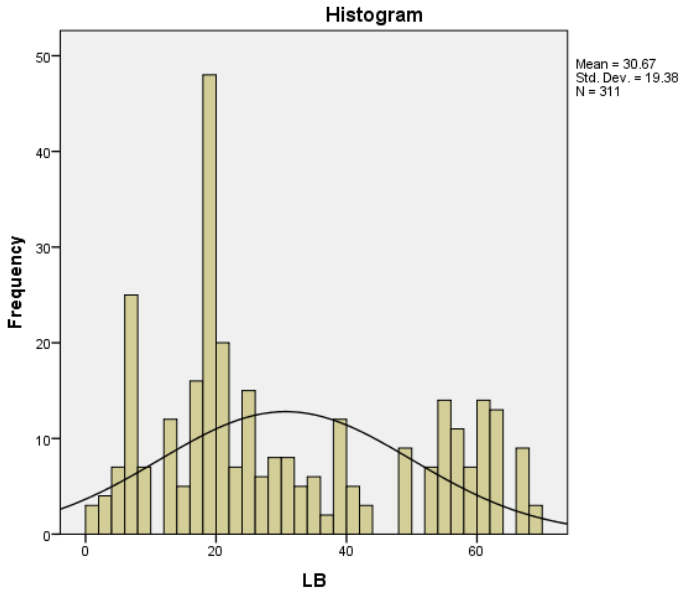


Figure 3.7. Response variable distribution.

Figure 3.7, shows the response variable returns of a skew of 0.496 with a 0.138 standard error.

This indicates asymmetric distribution.

### 3.6.3 Training Neural Network

The relationship between predicting and response variables was investigated using Multilayer Perceptron (MLP) network analysis method.

The input values  $X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14}$  shown in table 3.19 were identified as:

$X_1$  : ABC,  $X_2$  : D,  $X_3$  : C,  $X_4$  : AAC,  $X_5$  : A,  $X_6$  : VS,  $X_7$  : ABCT,  $X_8$  : DT,  $X_9$  : CT,  $X_{10}$  : AACT,  $X_{11}$  :AT,  $X_{12}$  : ASCORE,  $X_{13}$  : ATTEMPTS,  $X_{14}$  : AVG.

Considering the input values  $X_1, X_2, \dots, X_i$ , and the network weights were represented as  $W_1, W_2, \dots, W_i$  respectively,

The output  $\mathbf{X}_{15}$ :  $\mathbf{LS}$  is computed as:

$$y = \sum w_i \cdot x_i + bias.$$

The sum of input values multiplied by the corresponding weights gives the output. The access records collected from the learning contents of psychology form the input for this particular network. The encoded values from the sociology dataset form the desired outputs for the network. To generate the output, the product of weights and the corresponding inputs are computed. The results are added up and the output generated is compared to the expected until a threshold value is reached. The weights are modified to reduce the difference between the actual and expected output.

Bias shows the additional parameters for modifying the output together with the weighted sum of the inputs. The neural net function of r-language was used for building and training the network. The functions make it flexible to import datasets and do the training through a customized error and activation function.

This particular analysis used 311 rows of the dataset. The dataset was imported to a neural network and training was done by using the 14 input variables of the training set to predict the output class.

Using the neural net function of r language, the model randomly selected 0.7 training and 0.3 test sets from 311 rows of dataset as illustrated in the sample code below.

```
index = sample (1: nrow (data), round (0.70*nrow (data))).
```

```
train_data <- as.data.frame (data_scaled [index,]).
```

```
test_data <- as.data.frame (data_scaled [-index,]).
```

At each training stage, the neural network plot was generated as illustrated in the sample code below.

```
n = names (train_data).
```

```
f <- as.formula (paste ("LS ~", paste (n [! n %in% "LS"], collapse = "+ "))).
```

```
deep_net = neuralnet (f, data=train_data, hidden=c (5, 3, 2), linear.output=F, threshold = 0.001m,  
learningrate.factor = list (minus = 0.5, plus = 1.2), stepmax = 1e+05, rep = 10).
```

```
plot (deep_net, arrow.length = 0.15).
```

In the sample code above,

- f is the symbolic description of the model to be fitted in the sample code as shown above.
- data is the data frame containing the variables  $X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14}$ .
- the hidden function indicates the number of internal layers and nodes in the network. The nodes are represented using numbers.
- Threshold defines error function as a stopping criterion. The sample code above has a threshold value of 0.01
- The maximum training steps for a neural network are defined by the step max function. The training stops when the maximum step is reached.
- The number of training repetitions for the network is defined by the rep function. The repetitions in the sample code above are set to 10.

- The vector or lists for the multiplication factor for the network are defined by the learning rate. The vectors are defined in terms of upper and lower rates. The learning rate factor in the sample code above is set to 0.5 and 1.2 minimum and maximum.
- The network graph is displayed by the plot function.

The plotted network models for the different number of input variables are shown in Figures 3.8.1, 3.8.2, 3.8.3, and 3.8.4. The input layers are LS and CT variables generated from the records of interactions with the learning contents. The middle layers represent hidden neurons. The single output node is the response variable that is predicted. The black lines are the links between the nodes. Blue lines are bias added at each processing step.

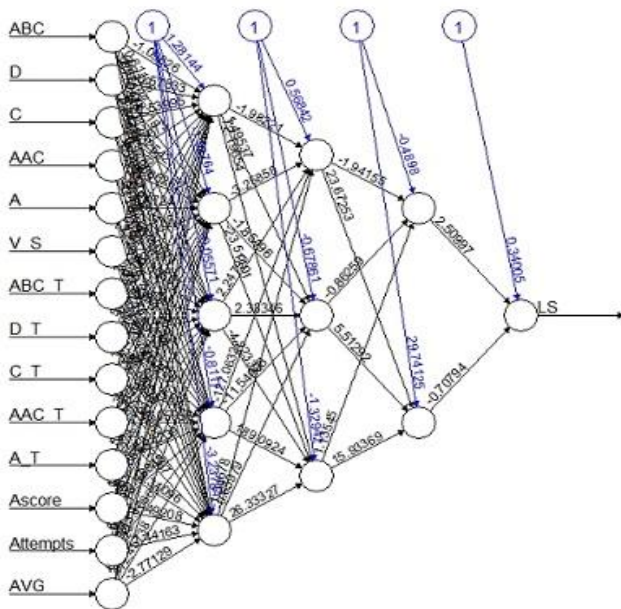
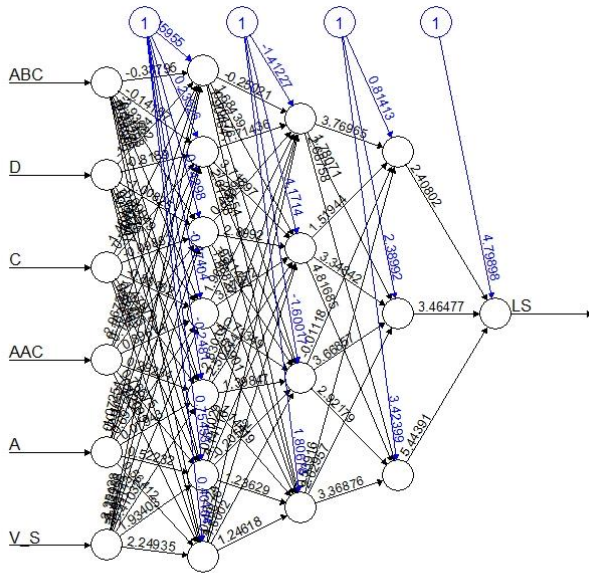


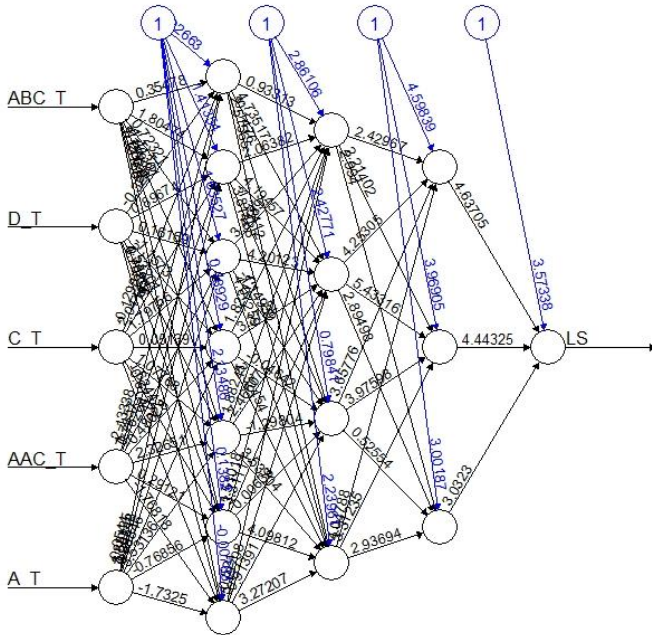
Figure 3.8.1. Neural Network Model Indicating all Input Variables.





Error: 134491.000705 Steps: 40

Figure 3.8.2. Neural Network Model Indicating LS Input Variables Related Content



Error: 126570.000938 Steps: 39

Figure 3.8.3. Neural Network Model Indicating LS Input Variables Related to Time

The variables indicate time spent on content pages.

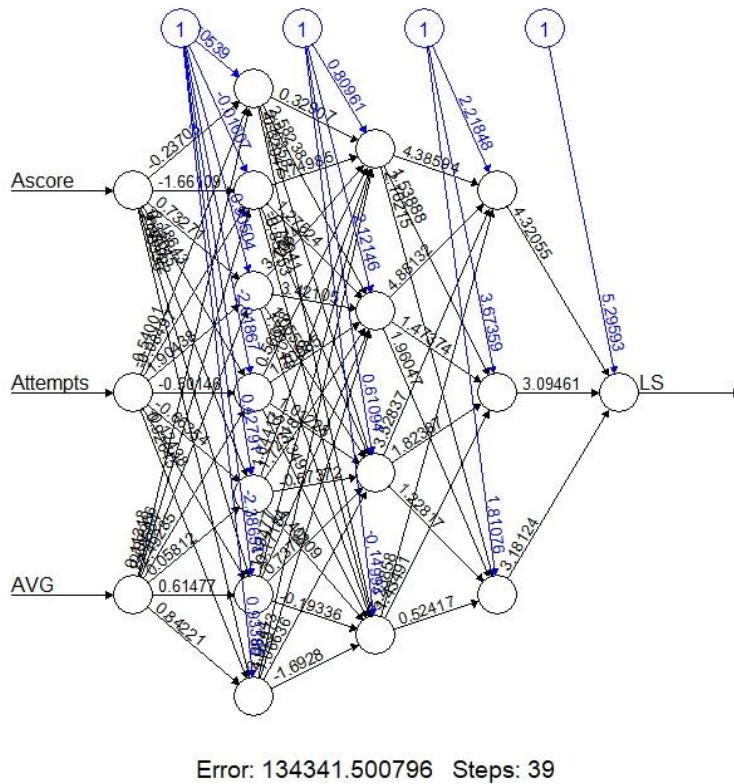


Figure 3.8.4. Neural Network Model Indicating Input Variables Related to CT

The dataset with input and output data was entered into the network. The output produced was checked against the expected results to measure the prediction accuracy of the network. The weights were repeatedly modified and subsequent pieces of training were conducted while increasing the number of hidden layers. At each training step, the performance of the model was assessed using the machine learning evaluation matrices.

### 3.6.4 Network Evaluation

The machine learning matrices commonly used to evaluate a machine model's performance were used in this study. These are accuracy, precision, recall, f-score, and area under the curve (AUC) (Zheng, 2015). Precision is an evaluation method used to measure the ratio of correctly predicted

positive observations to the total predicted positive observations as  $P = \frac{Tp}{Tp+Fp}$

Recall measures the ratio of correctly predicted positive observations to all observations in actual class as  $R = \frac{Tp}{Tp+Fn}$

F1-score is the weighted average of Precision and Recall as  $F1 = 2 * \frac{R*P}{R+P}$

Where  $T_P$ : True Positive,  $F_P$ : false positive,  $T_N$ : true negative,  $F_N$ : false negative.

Section 4.2 discusses evaluation results for learner behavior modeling using ANN.

### **3.7 Learner Behavior Modeling using Clustering Algorithms**

To compare the closeness of the results predicted by CogniLearn against an unsupervised machine learning method, the current study employed cluster algorithms to measure the similarity among the objects in the dataset collected from the LMS. The students with similar navigation patterns were expected to have high similarity. The students with different behavior were expected to have low similarity.

Clustering is an unsupervised technique that doesn't have a target column in the dataset and is suitable when nothing is known about the data to be analyzed. Clustering can be used to reduce the effect of outliers in the dataset (Rasyid and Andayani, 2018).

Clustering is an unsupervised machine-learning technique used to group objects which are closer to one another in a multidimensional feature space, usually to uncover some inherent structure that the data possesses (Han, Kamber, and Pei, 2012).

Clustering is commonly used in the analysis of learning management system data, to group together learners with similar browsing patterns and possibly share a common learning behavior (Poellhuber and Roy, 2019). Students can also be grouped according to their learning styles (Abdelhadi, Ibrahim, and Nurunnabi, 2019) and reading habits as proficient, struggling, and disengaged readers (Fang *et al.*, 2018). The results of cluster analysis can be evaluated using

internal or external properties by measuring the compactness, connectedness, and separation of the cluster partitions (Brock *et al.*, 2008).

Connectedness relates to what extent observations are placed in the same cluster as their nearest neighbors in the data space. Compactness assesses cluster similarity, usually by looking at the intra-cluster variance, while separation quantifies the degree of separation between clusters usually by measuring the distance between cluster centroids.

The Silhouette coefficient is one of the methods used to measure the compactness, connectedness, and separation of the cluster partitions. The Silhouette coefficient is calculated using the mean intra-cluster distance and the mean nearest-cluster distance for each sample.

For observation  $i$ , it is defined as:

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)},$$
 where  $a_i$  is the average distance between  $i$  and all other observations in the

same cluster, and  $b_i$  is the average distance between  $i$  and the observations in the nearest neighboring cluster. The silhouette width thus lies in the interval  $[-1; 1]$  and should be maximized.

The best value is 1 and the worst value is -1. Values near 0 indicate overlapping clusters. Negative values generally indicate that a sample has been assigned to the wrong cluster. The average silhouette approach measures the quality of a clustering. That is, it determines how well each object lies within its cluster. A high average silhouette width indicates good clustering.

The four algorithms from the base distribution of r- studio were used in this study. The algorithms are part of the add-on packages of the r-studio cluster function (Brock *et al.*, 2008). These algorithms are k-means, hierarchical, PAM, and Diana.

The k-means algorithm finds a specified number of clusters. Each cluster is represented by a centroid. Each point is assigned the closest centroid when several centroids are selected. The within-class sum of squares for a given number of clusters is kept minimal. (Hartigan and Wong, 2009). The k-means algorithm is available in the base distribution of R in the function `k-means()`. The hierarchical algorithm repeatedly puts together the two closest clusters until a single cluster is obtained (Kaufman and Rousseeuw, 2005). Each instance is placed in its group. The groups are successively joined following the closeness order. A dissimilarity matrix is used to estimate the closeness of the two clusters. This algorithm is implemented in the function `hclust()` of the R base distribution.

Divisive Analysis (DIANA) is a hierarchical clustering algorithm that constructs the hierarchy in the inverse order. At each subsequent step, the largest available cluster is split into two clusters until finally all clusters are comprised of single objects (Patnaik, Bhuyan, and Krishna Rao, 2016).

DIANA clustering algorithm is available in the base distribution of R in function `diana()`.

The partition around Medoids (PAM) algorithm is intended to find a sequence of objects called medoids that are centrally located in clusters. Objects that are tentatively defined as medoids are placed into a set  $S$  of a selected object. The algorithm is based on the search for  $k$  representative objects or medoids among the observations of the dataset. After finding a set of  $k$  medoids, clusters are constructed by assigning each observation to the nearest medoid. The goal of the algorithm is to minimize the average dissimilarity of objects to their closest selected object (Kaufman and Rousseeuw, 2005). PAM clustering algorithm is available in the base distribution of R in function `pam()`. In this study, silhouette analysis was used to assess the computational efficiency, accuracy, and quality of clusters created. The dataset provided in table 3.8 section 3.4.1 was used to analyze

the records of 389 students. Cluster analysis was done using the R software 3.5.1. The dataset was analyzed in three groups as shown:

- a. Variables related to LS – these are introductions, definitions, topics and sub-topics, summaries, activities, and illustrations.
- b. Variables related to CT – These are grades scored on tests, attempts on tests, and average percentage content coverage.
- c. Hybrid – all variables are considered.

In the current study, data exploration was done using the Euclidean distance method. Then the k-means, hierarchical, PAM, and Diana algorithms were used to create clusters based on Euclidean distances. Finally, the silhouette analysis was used to assess the computational efficiency, accuracy, and quality of clusters created. The next sections discuss the steps in detail.

### 3.7.1 Formulating Euclidean Distance

The study, first, explored the data using the Euclidean distance method. The Euclidean distance is a mathematical technique for computing the distance between the differences between a pair of objects. The method is used to estimate the square root distance between a pair of objects as:

$$d = \sqrt{\sum_{k=1}^n (x_{i_k} - x_{j_k})^2}$$

Where  $x_{ik}$ ,  $x_{jk}$  are coordinate points.

Table 3.10. Computing Euclidean Distance Between a Pair of Students

Student	ABC	D	C	AAC	A	V_T	ABC_T	D_T	C_T	AAC_T	AT	V_T	ascore	attempts	AVG
A	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$
B	$x'_1$	$x'_2$	$x'_3$	$x'_4$	$x'_5$	$x'_6$	$x'_7$	$x'_8$	$x'_9$	$x'_{10}$	$x'_{11}$	$x'_{12}$	$x'_{13}$	$x'_{14}$	$x'_{15}$

From Table 3.10 mapping the above formula to variables in the dataset, the coordinates  $X_1 \dots X_{15}$  are navigation patterns for student A and  $X'_1 \dots X'_{15}$  student B. The Euclidean distance between the two students is given by:

$$d_{AB} = \sqrt{(x_1 - x'_1)^2 + \dots + (x_{15} - x'_{15})^2}$$

Where  $X_1 \dots X_{14}$  are the variables shown in Table 3.10 above. The Euclidean distances between different pairs of students are repeatedly computed until the dataset is exhausted. The Euclidean distance between different coordinates was computed using the R base `dist()` function proposed by Kassambara (2015):

```
dist.eucl <- dist(LB, method = "euclidean").
```

```
round(as.matrix(dist.eucl)[1:10, 1:10], 1).
```

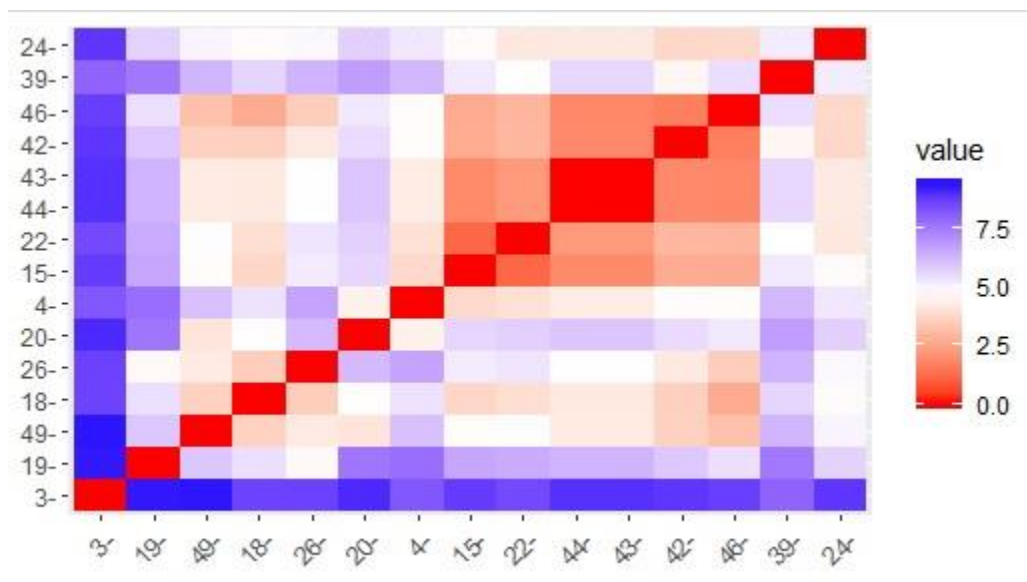
Where `LB` is a learning management system dataset file imported into `dist()` function. This process was important for estimating the similarity distance in content navigation shown by various students. The Euclidean distance facilitated the creation of learner profile clusters. Figure 3.9 shows the sample matrix generated by the R base `dist()` function. The subset contains the first 10 columns and rows. The numerical values are rounded off to the nearest ten.

	15	39	20	42	44	3	24	49	46	19
15	0.0	5.2	5.6	2.7	1.9	8.7	4.7	4.7	2.7	6.6
39	5.2	0.0	6.8	4.6	5.5	8.0	5.1	6.3	5.4	7.5
20	5.6	6.8	0.0	5.5	5.9	9.1	5.7	4.1	5.2	7.6
42	2.7	4.6	5.5	0.0	1.9	8.8	3.8	3.6	1.7	5.9
44	1.9	5.5	5.9	1.9	0.0	9.0	4.2	4.3	1.9	6.3
3	8.7	8.0	9.1	8.8	9.0	0.0	8.8	9.3	8.7	9.3
24	4.7	5.1	5.7	3.8	4.2	8.8	0.0	5.0	3.8	5.7
49	4.7	6.3	4.1	3.6	4.3	9.3	5.0	0.0	3.2	5.9
46	2.7	5.4	5.2	1.7	1.9	8.7	3.8	3.2	0.0	5.4
19	6.6	7.5	7.6	5.9	6.3	9.3	5.7	5.9	5.4	0.0

Figure 3.9. The Euclidean distance the matrix generated.

The values in Figure 3.9 represent the distances between students. In Figure 3.10, the values on the diagonal of the matrix represent the distance between students and themselves (which are zero). The distance matrices shown were visualized using the Visual Assessment of Cluster Tendency (VAT) algorithm using the R base `fviz_dist()` function.

*fviz\_dist(dist.eucl)*



*Figure 3.10. The visual assessment of the distance matrix*

The color level is proportional to the value of the dissimilarity between observations. Pure red if  $\text{dist}(x_i, x_j) = 0$  and pure blue corresponds to the highest value of Euclidean distance computed. The objects belonging to the same cluster are displayed in consecutive order. Thus students with similar browsing behavior are put in the same groups.



### 3.7.2 Cluster Analysis and Evaluation

To generate clusters for students based on the CT only, the variables considered were the marks scored (ascore), number of attempts on the tests done (attempts), and average percentage content coverage (AVG). Table 3.25 shows the results for the four clustering algorithms.

Clusters for students based only on the LS were generated using course outlines and introductions (ABC), definitions (D), topics and subtopics (C), summaries and end-of-unit tests (AAC), activities (A), illustrations (V\_S), time spent on outlines and introductions (ABC\_T), time spent on definitions (D\_T), time spent on topics and subtopics (C\_T), time spent on summaries and end of unit tests (AAC\_T), time spent on activity questions (A\_T), time spent on illustrations (V\_T).

The clusters for students based on hybrid behavior (CT and LS) were generated using all the variables. The variables were course outlines and introductions (ABC), definitions (D), topics and subtopics (C), summaries and end-of-unit tests (AAC), activities (A), illustrations (V\_S), time spent on outlines and introductions (ABC\_T), time spent on definitions (D\_T), time spent on topics and subtopics (C\_T), time spent on summaries and end of unit tests (AAC\_T), time spent on activity questions (A\_T), time spent on illustrations (V\_T), average marks scored on the tests (ascore), the average number of attempts on tests (attempts), average percentage content coverage (AVG).

In both cases, the dataset was used to create clusters between three and ten. The average silhouette was recorded for each number of clusters. The study used four algorithms to carry out an analysis of the dataset. The algorithms were k-means, pam, Diana, and hierarchical clustering. The results were evaluated using silhouette analysis.

Depending on the number of clusters required, the average silhouette width was recorded for each iteration. The clusters were computed using the `eclust()` function of R. The silhouette coefficient of observations was computed using the function `silhouette()` of the R package. The function returned the cluster number for each observation, neighbor, and silhouette width. The sample code below shows cluster analysis with k-means, PAM, and hierarchical cluster algorithms.

```
LB<-data.
```

```
kmeans <- eclust(LB, "kmeans", k = 3, ...)
```

```
pam <- eclust(LB, "pam", k = 3, ...)
```

```
hc. <- eclust(LB, "hclust", k = 3 ...)
```

```
diana <- eclust(LB, "diana", k = 3 ...)
```

From the sample code shown, LB is the variable for storing the dataset extracted from learning management system records. The dataset is imported into the `eclust()` function of R. The clustering algorithm is specified as k-means, Diana, pam, or hclust. The number of clusters to be generated is indicated by a numerical value. The results from the cluster analysis were evaluated using silhouette analysis. The `fviz_silhouette` function was used to compute the silhouette value for each clustering algorithm.

```
fviz_silhouette(kmeans,..)
```

```
fviz_silhouette(pam,...)
```

```
fviz_silhouette(hc,...)
```

```
fviz_silhouette(diana,...)
```

From the sample code shown, the `fviz_silhouette()` function returns the cluster number for each observation, neighbor, and silhouette width generated by each algorithm.

Section 4.3 discusses the evaluation results for learner behavior modeling using cluster analysis.

### 3.7.3 Incorporating Clusters in the CogniLearn Model

This study implemented a nearest-neighbor model to match the learner preferences to the existing groups in the system.

The current study assumes that students belonging to the same clusters share similar characteristics. This can be represented using IF <premises> THEN <conclusion>.as:

*IF:(1) Student A is in cluster 1 AND*

*(2) Students G and H are in cluster 1 AND*

*(3) Students G and H have unknown LS and CT*

*CONCLUSION: (4) Students G and H have the same LS and CT as A*

*THEN: content x is suitable for students G and H as well*

As shown in Table 3.11 below, the logic is that if students A, E, G, and H belong to the same cluster, but the LS and CT of G and H are unknown, then it can be concluded both G and H have similar LS and CT. Thus the system allocates student H contents similar to students A, E, and G.

Table 3.11. Allocating Content to Students with Unknown LS and CT

Student	cluster	content
A	1	X
B	2	Y
C	2	?
D	3	Z
E	1	X
F	3	?
G	1	?
H	1	?

As illustrated in appendix 14 where model interface and functionality are shown, students who are already in the system can access and view content matching their characteristics. However, there are a few cases where a student's LS and CT are not defined in the system. The only available information is the cluster a student belongs to.

To address this, a feature was added to the Cognilearn model to incorporate cluster modeling. Using this feature, a course lecturer can upload a cluster file with learners' profiles. The new students can access and see the cluster members with similar LS and CT. The system checks and allocates the new student contents similar to those in the same clusters.

### **3.8 Summary**

This chapter has described the design, development, and evaluation of three models used in the study. First, the dataset was extracted from the LMS and mapped to psychological variables matching the descriptions of FLSM and CTM. From this, a prototype estimating learning behavior based on the two theories was developed and evaluated with students in a classroom environment.

The study further explored other techniques suggested in the literature namely ANN and cluster analysis.

Concerning ANN, the dataset extracted from the LMS log was used to train the ANN. The ANN was trained to predict the LS and CT of each student based on FLSM and CTM. The input dataset was entered into the network and trained to classify learners according to LS and CT. Precision, recall, f-score, and area under the curve (AUC) were used to evaluate the prediction ability of the model.

Concerning cluster analysis, four algorithms were used to analyze the dataset in this study. These algorithms are k-means, hierarchical, PAM, and Diana. The dataset was analyzed in three groups namely variables related to LS, CT, and both. Finally, the silhouette analysis was used to assess the computational efficiency, accuracy, and quality of clusters created. The next chapter discusses the results.

## **CHAPTER 4: RESULTS AND DISCUSSIONS**

This study used three methods to identify LS and CT in an LMS. First, a prototype was designed to estimate LS and CT from records of learner interactions with an LMS. The prototype was evaluated with students using a case study approach. Second, a neural network model for predicting LS and CT was developed and evaluated. Finally, a cluster-based model was developed and evaluated.

This chapter discusses the evaluation results for the three approaches.

Section 4.1 discusses the evaluation results for the CogniLearn model. Section 4.2 discusses the evaluation results for the ANN. Section 4.3 discusses the evaluation results for the cluster analysis method. Finally, the results are compared with similar studies cited in the literature.

### **4.1 Results for Prototype Evaluation**

The model was evaluated using the results which were obtained from the administration of the manual psychometric tools to students in a classroom environment as discussed in section 3.5.

The results gathered using manual tools were compared against those generated by the model.

#### **4.1.1 Descriptive Statistics**

The survey for LS recorded a response rate of 62% (124 out of 200 ILS questionnaires were correctly filled and returned). The remaining 76 copies either wrongly filled or incomplete were not considered.

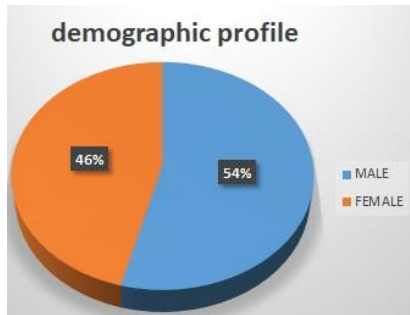


Figure 4.1. Distribution of Participants by Gender (N=124)

Male participants represented 54% (67) of the population with 46% (57) being female.

Table 4.1. Distribution of Learning Styles by Strengths

		Count	Layer Total N %
Active	None	65	52.4%
	Mild	41	33.1%
	Moderate	17	13.7%
	Strong	1	0.8%
Reflective	None	60	48.4%
	Mild	39	31.5%
	moderate	22	17.7%
	strong	3	2.4%
Sensing	none	42	33.9%
	Mild	38	30.6%
	moderate	38	30.6%
	strong	6	4.8%
Intuitive	none	83	66.9%
	Mild	24	19.4%
	moderate	10	8.1%
	strong	7	5.6%
Visual	none	25	20.2%
	Mild	32	25.8%
	moderate	45	36.3%
	strong	22	17.7%
Verbal	none	100	80.6%
	Mild	16	12.9%
	moderate	4	3.2%
	strong	4	3.2%
Sequential	none	44	35.5%
	Mild	44	35.5%
	moderate	29	23.4%
	strong	7	5.6%
Global	none	82	66.1%
	Mild	26	21.0%
	moderate	14	11.3%
	strong	2	1.6%

From table 4.1, 52.4% of the respondents showed no preference for active LS, while 33.1%, 13.7%, and 0.8% of the respondents showed a preference for mild, moderate, and strong preferences respectively. For reflective LS, 48.4% showed no preference; while 31.5%, 17.7%, and 2.4% of respondents indicated mild, moderate, and strong preferences respectively. 33.9% showed no preference for sensing LS while 30.6%, 30.6%, and 4.8% of respondents expressed mild, moderate, and strong preferences respectively. 66.9% showed no preference for intuitive LS while 19.4%, 8.1%, and 5.6% indicated mild, moderate, and strong preferences respectively. 20.2% showed no preference for visual LS while 25.8%, 36.3%, and 17.7% indicated mild, moderate, and strong preferences respectively. 80.6% showed no preference for verbal LS while 12.9%, 3.2%, and 3.2% expressed mild, moderate, and strong preferences respectively. 35.5% showed no preference for sequential LS while 35.5%, 23.4%, and 5.6% indicated their mild, moderate, and strong preferences. 66.1% showed no preference for global LS while 21.0%, 11.3%, and 1.6% denoted their mild, moderate, and strong preferences respectively.

The survey for CT recorded a response rate of 58.5%, 56.5%, 55%, and 57% for ALA, IRA, WMC, and IPS respectively. Tables 4.2, 4.3, 4.4, and 4.5 show the distribution of CT by strengths.

*Table 4.2. Distribution of ALA by Strengths*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low	17	14.5	14.5	14.5
	Moderate	52	44.4	44.4	59.0
	High	48	41.0	41.0	100.0
	Total	117	100.0	100.0	



117 out of 200 sample students attempted online cognitive tests related to ALA. The results indicated that 14.5%, 44.4%, and 41.0% of the respondents had low, moderate, and high ALA abilities respectively.

*Table 4.3. Distribution of IRA by Strengths*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low	39	34.5	34.5	34.5
	Moderate	51	45.1	45.1	79.6
	High	23	20.4	20.4	100.0
	Total	113	100.0	100.0	

113 out of 200 respondent students attempted online cognitive tests related to IRA. The results indicated that 34.5%, 45.1%, and 20.4% of the respondents had low, moderate, and high IRAs respectively. See table 4.3.

*Table 4.4. Distribution of WMC by Strengths*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low	9	8.2	8.2	8.2
	Moderate	37	33.6	33.6	41.8
	High	64	58.2	58.2	100.0
	Total	110	100.0	100.0	

110 out of 200 attempted online cognitive tests related to WMC. The results indicated that 8.2%, 33.6%, and 58.2% of the respondents had low, moderate, and high WMC respectively.

Table 4.5. Distribution of IPS by Strengths

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low	29	25.4	25.4	25.4
	Moderate	60	52.6	52.6	78.1
	High	25	21.9	21.9	100.0
	Total	114	100.0	100.0	

114 out of 200 subjects attempted online cognitive tests related to IPS. The results indicated that 25.4%, 52.6%, and 21.9% of the respondents had low, moderate, and high IPS respectively.

#### 4.1.2 Evaluation Results for LS Prediction

Based on the number of visits on pages matching the descriptions of active-reflective, sensing-intuitive, sequential-global LS, the prototype predicted the preferences for each student as low, moderate, and high. The interrater reliability analysis was done to compare the results of the ILS questionnaire against those predicted by the prototype. The interrater reliability analysis results for LS are shown in Tables 4.6 – 4.8.

##### *Evaluation for Results Active - Reflective LS*

Table 4.6 shows the active-reflective raters recorded Kappa = 0.260 (p <0.018). This can be interpreted as a fair agreement between the results predicted by the model for active–reflective against those gathered through the ILS questionnaire.

Table 4.6. Interrater Reliability Analysis Results for Active-Reflective LS

a. **Cross tabulation**

		model estimated_output		Total
		active	Reflective	
ILS	Active	15	19	34
	Reflective	4	22	26
Total		19	41	60

b. **Symmetric measures**

		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Measure of Agreement	Kappa	.269	.107	2.371	.018
N of Valid Cases		60			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

***Evaluation Results for Sensing – Intuitive LS***

Based on the number of visits to the learning objects matching the descriptions of sensing-intuitive LS, the prototype predicted the preferences for each student as low, moderate, and high. The interrater reliability analysis was done to compare the results ILS questionnaire and those predicted by the prototype. Table 4.7. below shows the sensing-intuitive raters recorded Kappa = 0.595 (p <0.002). This can be interpreted as a moderate agreement between the results predicted by the model for sensing-intuitive against those gathered through the ILS questionnaire.

Table 4.7. Interrater Reliability Analysis Results for Sensing – Intuitive LS

a. **Cross tabulation**

		model estimated_output		Total
		sensing	Intuitive	
ILS	Sensing	12	4	16
	Intuitive	1	8	9
Total		13	12	25

b. **Symmetric measures**

		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Measure of Agreement	Kappa	.595	.157	3.069	.002
N of Valid Cases		25			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

***Evaluation Results for Sequential – Global LS***

Using the number of visits to the learning objects matching the descriptions of sequential-global LS, the prototype predicted the preferences for each student as low, moderate, and high. The interrater reliability analysis was done to compare the results ILS questionnaire and those predicted by the prototype.

Table 4.8. below shows the sequential-global raters recorded Kappa = 0.326 ( $p < 0.004$ ). This can be interpreted as a moderate agreement between the results predicted by the model for sequential-global against those gathered through the ILS questionnaire.

*Table 4.8. Interrater Reliability Analysis Results for Sequential – Global LS*

**a. Cross tabulation**

		model estimated_output		Total
		Sequential	Global	
ILS	sequential	31	1	32
	Global	13	6	19
Total		44	7	51

**b. Symmetric measures**

	Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Measure of Agreement Kappa	.326	.122	2.855	.004
N of Valid Cases	51			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

The interrater reliability test was not conducted on the visual-verbal learning style. All the students who filled out the ILS questionnaire indicated preferences for visual content.

### 4.1.3 Evaluation Results for CT Prediction

The study used the performance of online tests as a psychometric measure of CT namely ALA, IRA, WMC, and IPS. Both the CT results predicted by the prototype and the performance results on psychometric tests were classified as low, moderate, and high. The interrater reliability analysis tests were done to compare the results of the psychometric tests against those generated by the model. The interrater reliability analysis results for CT are shown in Tables 4.9 – 4.12.

#### *Evaluation Results for Associative Learning Ability*

According to CTM, an associative learner reads by creating a connection between new and previously acquired knowledge. In this study, a student who frequently visited content pages that were already viewed was assumed to be an associative learner. Paired associate tests were administered to test the associative learning ability of the students. The paired-associate tests evaluate the ability to remember and recall specific events, paired with the context in which they occur. The results of paired-associate tests were classified as low, moderate, and high. Similarly, based on the number of revisited content pages, the prototype predicted associative learning ability as low, moderate, and high. The interrater reliability analysis was done to compare the results.

Table 4.9. shows that the raters for associative learning ability recorded Kappa = 0.012 ( $p < 0.640$ ). This can be interpreted as a slight agreement between the results predicted by the model against those gathered through the online cognitive tests.

*Table 4.9. Interrater Reliability Analysis Results for Associative Learning Ability.*

#### a. Cross tabulation

	model estimated_output		Total
	Low	Moderate	
Low	0	1	1
Moderate	2	4	6

Online High cognitive tests	25	20	45
Total	27	25	52

**b. Symmetric measures**

	Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Measure of Agreement Kappa N of Valid Cases	.012 52	.027	.467	.640

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

***Evaluation Results for Information Processing Speed***

According to CTM, information processing speed checks to what extent a learner can quickly retrieve the knowledge from memory and use the information to make the correct decision. In this study, a student who made a few attempts on an online test but scored a good grade was assumed to have a high information processing speed. A student who made multiple attempts with a score below average was considered to have a low working memory capacity. Mental speed tests were used to test the information processing speed. The mental speed tests measured how quickly a student could process information and make decisions based on the available information. The results of mental speed tests were classified as low, moderate, and high. Based on the number of attempts and grades on online quizzes administered during coursework, the prototype predicted information processing speed as low, moderate, and high. The interrater reliability analysis was done to compare the results of abstract reasoning tests and those predicted by the prototype.

Table 4.10. illustrates the raters for information processing speed recorded Kappa = 0.455 ( $p < 0.067$ ). This can be interpreted as a moderate agreement between the results predicted by the model against those gathered through the online cognitive tests.

Table 4.10. Interrater Reliability Analysis Results for Information Processing Speed

a. **Cross tabulation results**

		Model estimated_output		Total
		moderate	High	
Online cognitive tests	Moderate	4	3	7
	High	1	7	8
Total		5	10	15

b. **Symmetric measures**

		Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Measure of Agreement	Kappa	.455	.224	1.830	.067
N of Valid Cases		15			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

***Evaluation Results for Inductive Reasoning Ability***

According to CTM, inductive reasoning checks to what extent a student can make concepts from examples. In this study, a student who frequently visited pages with contents on examples was assumed to have high inductive reasoning. Those who made fewer visits or none were assumed to have low inductive reasoning. Abstract reasoning tests were used to test the inductive reasoning ability of the students. The abstract reasoning tests measured the cognitive ability to apply rules to information to arrive at a logical conclusion. The results of abstract reasoning tests were classified as low, moderate, and high. Based on visits to content pages with examples, the prototype predicted

inductive reasoning ability as low, moderate, and high. The interrater reliability analysis was done to compare the results of abstract reasoning tests and those predicted by the prototype.

As shown in Table 4.11., the raters for inductive reasoning ability recorded Kappa = 0.224 ( $p < 0.052$ ). This can be interpreted as a fair agreement between the results predicted by the model against those gathered through the online cognitive tests.

Table 4.11. Interrater Reliability Analysis Results for Inductive Reasoning Ability

a. **Cross tabulation results**

		model estimated_output			Total
		Low	moderate	High	
Online cognitive tests	Low	2	0	0	2
	Moderate	3	3	0	6
	High	5	1	1	7
Total		10	4	1	15

b. **Symmetric Measures**

	Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Measure of Agreement Kappa	.224	.117	1.944	.052
N of Valid Cases	15			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

**Evaluation Results for Working Memory Capacity**

According to CTM, working memory capacity checks to what extent a student can concentrate on a reading task without losing focus. In this study, frequently visiting content pages by navigating forward was assumed to be a sign of high working memory capacity. Spatial span tests were used to test the working memory capacity of the students. The spatial span tests measured the ability to remember the relationships between objects in space. The results of spatial span tests were classified as low, moderate, and high. Based on serial navigation on content pages, the prototype



predicted working memory capacity as low, moderate, and high. The interrater reliability analysis was done to compare the results.

Table 4.12 shows the raters for working memory capacity recorded Kappa = -0.087 ( $p < 0.561$ ). This can be interpreted as a disagreement between the results predicted by the model against those gathered through the online cognitive tests.

Table 4.12. Interrater Reliability Analysis Results for Working Memory Capacity

a. **Cross tabulation**

	model estimated_output			Total
	Low	Moderate	high	
Online low cognitive tests	0	0	1	1
high	2	2	10	14
Total	2	2	11	15

b. **Symmetric Measures**

	Value	Asymptotic Standardized Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Measure of Agreement Kappa	-.087	.074	-.582	.561
N of Valid Cases	15			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

## 4.2 Evaluation Results for the Neural Network Model

As stated in section 3.4, the machine learning matrices commonly used to evaluate a machine model performance we used to assess the prediction accuracy of ANN concerning learner behavior modeling in an LMS. The matrices used are accuracy, precision, recall, f-score, and area under the curve. This section discusses evaluation results. As discussed in section 3.4.4, precision is an evaluation method used to measure the ratio of correctly predicted positive observations to the

total predicted positive observations. Recall measures the ratio of correctly predicted positive observations to all observations in an actual class. F1-score is the weighted average of Precision and Recall.

The dataset was analyzed in three groups as shown:

- d. Hybrid – all variables are considered.
- e. Variables related to LS – these are introductions, definitions, topics and subtopics, summaries, activities, and illustrations.
- f. Variables related to CT – These are grades scored on tests, attempts on tests, and average percentage content coverage.

First, the variables related to LS and CT were analyzed together. Next, the variables related to LS and CT were each analyzed separately.

#### **4.2.1 Evaluation Results for all variables**

To predict CT and LS, all the variables were considered for training the ANN. The variables were course outlines and introductions (ABC), definitions (D), topics and sub-topics (C), summaries and end-of-unit tests (AAC), activities (A), illustrations (V\_S), time spent on outlines and introductions (ABC\_T), time spent on definitions (D\_T), time spent on topics and sub-topics (C\_T), time spent on summaries and end of unit tests (AAC\_T), time spent on activity questions (A\_T), time spent on illustrations (V\_T), average marks scored on the tests (ascore), the average number of attempts on tests (attempts), average percentage content coverage (AVG).

The best prediction results were returned by a network with combinations of 5-3-2 hidden nodes. Figure 4.2 shows the confusion matrix table for a 5-3-2 network. There are 32 cases predicted as

correct and correct. 27 cases were predicted as incorrect and incorrect. There are 18 cases predicted as correct but incorrect, while 16 cases predicted as incorrect were correct.

The accuracy, precision, recall, f-score, and kappa results for the network were 0.63, 0.67, 0.64, 0.65, and 0.26 respectively. The kappa results are indications that the collected data moderately represents the LS and CT variables investigated. Figure 4.3 shows the plot for the true positive rate (TPR) against the false positive rate (FPR). It can be seen that an increase in TPR also increases FPR. The model returns AUC results of 0.64.

```

Confusion Matrix and Statistics

          Reference
Prediction 0 1
0          32 16
1          18 27

    Accuracy : 0.6344086
    95% CI   : (0.528122, 0.7318809)
  No Information Rate : 0.5376344
    P-Value [Acc > NIR] : 0.0377874

    Kappa : 0.2670376
  Mcnemar's Test P-Value : 0.8638317

    Precision : 0.6666667
    Recall    : 0.6400000
     F1      : 0.6530612
  Prevalence : 0.5376344
  Detection Rate : 0.3440860
  Detection Prevalence : 0.5161290
  Balanced Accuracy : 0.6339535

  'Positive' Class : 0
  
```

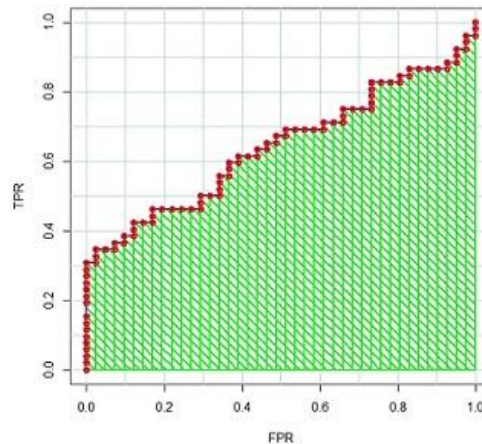


Figure 4.2: Confusion matrix table.

Figure 4.3: Area Under the Curve (AUC)

Sample prediction results for an ANN model are shown in appendix 33. The numbers in the first column stand for students. The values in the second column indicate the expected results. The values in the third column indicate the predicted results by the model. The number 0 stands for false and 1 for true. For example, the expected results for student number 1 are false. The prediction results indicate false. The prediction is therefore correct. The expected results for student number 5 are true. The prediction results indicate true. The prediction is therefore correct. The expected

results for student number 3 are true. The prediction returns false results. The prediction is therefore incorrect.

#### 4.2.2 Evaluation Results Variables Related to LS

To predict LS only, the variables considered were course outlines and introductions (ABC), definitions (D), topics and sub-topics (C), summaries and end-of-unit tests (AAC), activities (A), illustrations (V\_S), time spent on outlines and introductions (ABC\_T), time spent on definitions (D\_T), time spent on topics and sub-topics (C\_T), time spent on summaries and end of unit tests (AAC\_T), time spent on activity questions (A\_T), time spent on illustrations (V\_T). These variables related to LS were first analyzed. Then the analysis of variables related to the content pages visited was done. Finally, the variables related to the time spent on the content pages were analyzed. Tables 4.13, 4.14, 4.15, and 4.16 show the training results for different network layers.

*Table 4.13: Evaluation Results for all LS Variables*

Hidden	Precision	Recall	f-score	AUC
1	0.46	0.48	0.47	0.567
2	0.67	0.5	0.57	0.544
3	0.54	0.61	0.458	0.506
1,1	0.44	0.47	0.459	0.561
2,2	0.45	0.46	0.457	0.553
3,3	0.49	0.49	0.49	0.462
4,4	0.62	0.56	0.59	0.478
5,5	0.49	0.45	0.47	0.517
5,3,2	0.44	0.43	0.43	0.559
6,3,2	0.44	0.55	0.49	0.635
7,4,3	0.508	0.508	0.508	0.589
7,4,3,1	0.537	0.537	0.537	0.571

As shown in Table 4.13 above, when all the variables related to LS were used, the precision, recall, and f-score results for a single layer configuration range between 0.55 and 0.48 for maximum and minimum respectively. Two-layer configurations range between 0.56 and 0.47 for maximum and minimum respectively. Three-layer configurations range between 0.63 and 0.59 for maximum and minimum respectively. There were no significant changes recorded above the three-layer configurations. The ANN model for all variables related to LS is shown in Figure 3.8.1.

*Table 4.14. Evaluation Results for LS Variables Related to Content Pages Visited*

Hidden	Precision	Recall	f-score	AUC
1	0.547	0.51	0.52	0.507
2	0.58	0.60	0.59	0.59
3	0.52	0.50	0.51	0.528
1,1	0.63	0.56	0.60	0.55
2,2	0.56	0.51	0.53	0.56
3,3	0.39	0.50	0.43	0.63
3,3	0.42	0.43	0.43	0.61
4,4	0.40	0.50	0.44	0.52
5,5	0.56	0.45	0.50	0.50
5,3,2	0.67	0.64	0.65	0.64
6,3,2	0.58	0.659	0.620	0.64
7,4,3	0.61	0.70	0.65	0.47
7,4,3,1	0.51	0.67	0.58	0.49

From Table 4.14 above, when all the LS variables related to content pages visited were analyzed the precision, recall, and f-score results for a single layer configuration range between 0.60 and 0.48 for maximum and minimum respectively. Two-layer configurations range between 0.70 and 0.58 for maximum and minimum respectively. Three-layer configurations range between 0.63 and 0.59 for maximum and minimum respectively. There were no significant changes recorded above the three-layer configurations. The ANN model for all variables related to LS is shown in Figure

3.8.2.

*Table 4.15. Evaluation Results for LS Variables Related to Time Spent on the Content Pages*

Hidden	Precision	Recall	f-score	AUC
1	0.5	0.55	0.524	0.483
2	0.39	0.44	0.42	0.526
3	0.45	0.54	0.45	0.488
1,1	0.55	0.73	0.63	0.559
2,2	0.60	0.54	0.57	0.553
3,3	0.47	0.53	0.50	0.550
4,4	0.52	0.59	0.55	0.511
5,5	0.56	0.47	0.51	0.511
5,3,2	0.55	0.47	0.50	0.560
6,3,2	0.53	0.48	0.50	0.509
7,4,3	0.59	0.5	0.54	0.524
7,4,3,1	0.45	0.4	0.42	0.459

From Table 4.15 above, when all the LS variables related to time spent on pages visited were analyzed; the precision, recall, and f-score results for a single layer configuration range between 0.54 and 0.39 for maximum and minimum respectively. Two-layer configurations range between 0.73 and 0.47 for maximum and minimum respectively. Three-layer configurations range between 0.59 and 0.47 for maximum and minimum respectively. There were no significant changes recorded above the three-layer configurations. The ANN model for all variables related to LS is shown in figure 3.8.3.

### 4.2.3 Evaluation Results for Variables Related to CT

To predict CT only, the variables considered were the marks scored (ascore), number of attempts on the tests done (attempts), and average percentage content coverage (AVG). Table 4.16 shows the training results for different network layers.

*Table 4.16: Evaluation Results for Variables Related to Cognitive traits*

Hidden	Precision	Recall	f-score	AUC
1	0.42	0.44	0.43	0.535
2	0.56	0.56	0.56	0.586
3	0.6	0.67	0.63	0.533
1,1	0.37	0.43	0.4	0.605
2,2	0.49	0.5	0.49	0.47
3,3	0.45	0.57	0.50	0.508
4,4	0.6	0.5	0.56	0.574
5,5	0.5	0.52	0.514	0.581
5,3,2	0.37	0.57	0.45	0.592
6,3,2	0.557	0.517	0.513	0.621
7,4,3	0.44	0.47	0.45	0.624
7,4,3,1	0.37	0.47	0.42	0.550

From Table 4.16 above, it can be observed that when only variables related to CT were used, the precision, recall, and f-score results for a single layer configuration range between 0.67 and 0.43 for maximum and minimum respectively. Two-layer configurations range between 0.60 and 0.37 for maximum and minimum respectively. Three-layer configurations range between 0.55 and 0.37

for maximum and minimum respectively. There are no significant changes recorded above the three-layer configurations. The ANN model for all variables related to CT is shown in Figure 3.8.4.

The summary of results for all variables indicating the minimum and maximum for each criterion described above are shown in table 4.17 below.

*Table 4.17: Summary of Evaluation Results*

Variables	layers	Precision		Recall		f-score		AUC	
		Max	min	max	min	max	min	max	min
LS time and content	1	0.67	0.45	0.61	0.48	0.57	0.45	0.56	0.50
	2	0.62	0.44	0.56	0.46	0.59	0.45	0.56	0.46
	3	0.53	0.44	0.55	0.43	0.53	0.43	0.63	0.55
LS content	1	0.68	0.52	0.60	0.50	0.59	0.51	0.59	0.50
	2	0.63	0.39	0.56	0.42	0.60	0.43	0.63	0.50
	3	0.67	0.58	0.70	0.64	0.65	0.62	0.64	0.47
LS time	1	0.50	0.39	0.55	0.44	0.52	0.42	0.52	0.48
	2	0.60	0.47	0.73	0.47	0.63	0.50	0.55	0.51
	3	0.59	0.53	0.50	0.47	0.54	0.50	0.56	0.50
CT	1	0.60	0.42	0.67	0.44	0.63	0.43	0.58	0.53
	2	0.60	0.37	0.57	0.43	0.56	0.40	0.60	0.47
	3	0.50	0.39	0.57	0.47	0.51	0.45	0.62	0.59

From Table 4.17 above, when only variables related to LS (learning contents viewed and time spent) were used, the minimum and maximum precision, recall, f-score, and AUC for a single layer range between 0.45 and 0.67, 0.48 and 0.61, 0.45 and 0.57, 0.50 and 0.56 respectively. The minimum and maximum precision, recall, f-score, and AUC for two layers range between 0.44 and 0.62, 0.46 and 0.56, 0.45 and 0.59, 0.46 and 0.56 respectively. The minimum and maximum precision, recall, f-score, and AUC for three layers range between 0.44 and 0.53, 0.43 and 0.55, 0.43 and 0.53, 0.55 and 0.63 respectively.

When only variables related to LS (learning contents viewed) were used, the minimum and maximum precision, recall, f-score, and AUC for a single layer range between 0.52 and 0.68, 0.50 and 0.60, 0.51 and 0.59, 0.50 and 0.59 respectively. The minimum and maximum precision, recall,

f-score, and AUC for two layers range between 0.39 and 0.63, 0.42 and 0.56, 0.43 and 0.60, 0.50 and 0.63 respectively. The minimum and maximum precision, recall, f-score, and AUC for three layers range between 0.58 and 0.67, 0.64 and 0.70, 0.62 and 0.65, 0.47 and 0.64 respectively.

When only variables related to LS (time spent on content) were used, the minimum and maximum precision, recall, f-score, and AUC for a single layer range between 0.39 and 0.50, 0.44 and 0.55, 0.42 and 0.52, 0.48 and 0.52 respectively. The minimum and maximum precision, recall, f-score, and AUC for two layers range between 0.47 and 0.6, 0.44 and 0.67, 0.43 and 0.63, 0.53 and 0.58 respectively. The minimum and maximum precision, recall, f-score, and AUC for three layers range between 0.53 and 0.59, 0.47 and 0.50, 0.50 and 0.54, 0.50 and 0.56 respectively.

When only variables related to CT were used, the minimum and maximum precision, recall, f-score, and AUC for a single layer range between 0.42 and 0.60, 0.44 and 0.67, 0.43 and 0.63, 0.53 and 0.58 respectively. The minimum and maximum precision, recall, f-score, and AUC for two layers range between 0.37 and 0.60, 0.43 and 0.57, 0.40 and 0.56, 0.47 and 0.60 respectively. The minimum and maximum precision, recall, f-score, and AUC for three layers range between 0.39 and 0.50, 0.47 and 0.57, 0.45 and 0.51, 0.59 and 0.52 respectively.

### **4.3 Evaluation Results for Cluster Models**

As stated in section 3.5, four algorithms from the base distribution of the R-studio were used in this study. These algorithms are k-means, hierarchical, PAM, and Diana. The silhouette analysis was used to assess the computational efficiency, accuracy, and quality of clusters created. This section presents the results of the cluster analysis evaluation. The first analysis examined the impact of time spent on content, and access frequency on the cluster quality by considering variables related to CT. The second analysis examined the impact of time spent on content on the



cluster quality by considering variables related to time spent on contents matching LS. Finally, the last analysis examined the impact of content frequency on the cluster quality by considering variables to the number of contents viewed that matched LS. The dataset was used to create clusters between three and ten. The average silhouette was recorded for each number of clusters.

### 4.3.1 Results for Variables Related to CT

To generate clusters for students based on cognitive traits, only the datasets indicating variables for cognitive traits were considered. The variables considered were the marks scored (ascore), number of attempts on the tests done (attempts), and average percentage content coverage (AVG). The dataset was used to create clusters between three and ten. The average silhouette width was recorded for each number of clusters. Table 4.18 below presents the results for the four clustering algorithms.

*Table 4.18. Silhouette Analysis Results for Variables Related to Cognitive Traits*

CLUSTERS	Average silhouette width CT variables only			
	k-means	PAM	Diana	HC
3	0.5	0.52	0.46	0.52
4	0.54	0.57	0.46	0.584
5	0.63	0.65	0.47	0.62
6	0.65	0.71	0.48	0.64
7	0.67	0.73	0.52	0.68
8	0.73	0.78	0.59	0.77
9	0.78	0.82	0.59	0.81
10	0.81	0.85	0.57	0.83

It is apparent from Table 4.18 above that the compactness, connectedness, and separation were getting better in all algorithms as the size of the clusters increased. A high average silhouette width of 0.81, 0.85, and 0.83 were observed on k-means, PAM, and HC respectively indicating good evaluation results. Interestingly, the Diana algorithm generated lower silhouette widths with all cluster sizes compared to the rest. Figure 4.4 below shows the average silhouette plot for CT variables based on the PAM cluster algorithm. The column cluster indicates groups. The cluster

has 10 groups. The column size indicates the number of students placed in each group. The last column indicates the average silhouette width for each group.

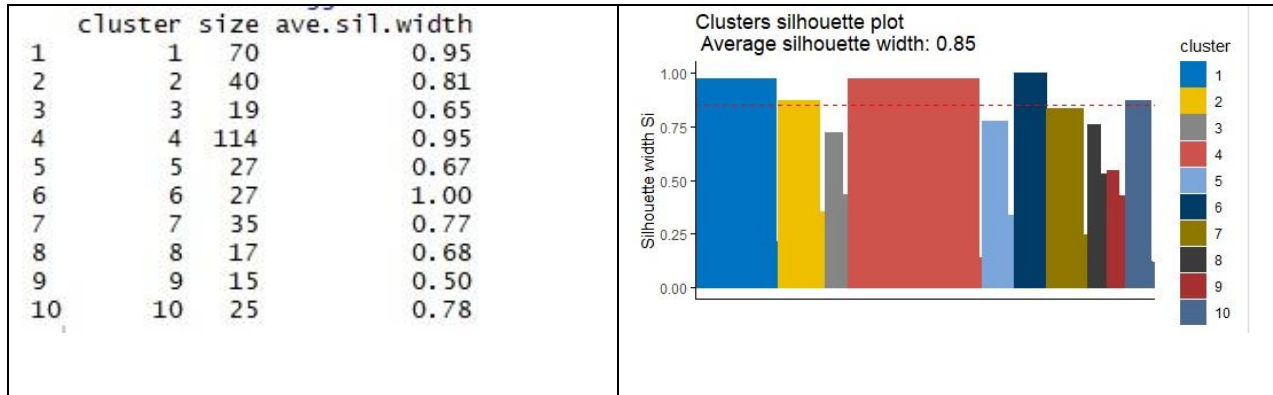


Figure 4.4. Silhouette Plot for Variables Related to CT (  $k=10$  )

From figure 4.4, it can be seen that the strongest evidence of the average silhouette width was recorded in cluster 6 with 27 students. The maximum number of students was contained in cluster 4. Cluster 9 had the least number of students. Overall, the average silhouette width was 0.85. This analysis revealed that the compactness, connectedness, and separation of the cluster partitions were high when only CT variables were used.

### 4.3.2 Results for Variables Related to LS

Table 4.19 shows cluster analysis for variables related to LS. The first analysis generated clusters for the frequency of content access and time by considering all variables related to LS. The second analysis examined the impact of time spent on content on the cluster quality by considering only variables related to time spent on contents matching LS. Finally, the last analysis examined the impact of content visit frequency on the cluster quality by considering only variables related to content viewed matching LS.

Table 4.19: Cluster analysis results for variables related to LS

	% content access and time spent variables	time spent on content variables only	% content access variables only
--	-------------------------------------------	--------------------------------------	---------------------------------

CLUSTERS	k-means	PAM	Diana	HC		k-means	PAM	Diana	HC		k-means	PAM	Diana	HC
3	0.22	0.17	0.23	0.2		0.23	0.23	0.2	0.18		0.36	0.36	0.4	0.3
4	0.22	0.19	0.22	0.2		0.23	0.14	0.17	0.2		0.35	0.36	0.37	0.4
5	0.21	0.18	0.21	0.2		0.24	0.24	0.2	0.22		0.36	0.37	0.35	0.4
6	0.22	0.19	0.22	0.2		0.22	0.22	0.2	0.23		0.4	0.39	0.35	0.4
7	0.22	0.19	0.21	0.2		0.23	0.2	0.19	0.24		0.41	0.4	0.35	0.4
8	0.23	0.16	0.21	0.21		0.23	0.21	0.2	0.18		0.42	0.37	0.35	0.4
9	0.21	0.17	0.2	0.21		0.24	0.19	0.2	0.2		0.42	0.39	0.32	0.4
10	0.22	0.19	0.21	0.18		0.26	0.19	0.2	0.21		0.44	0.4	0.33	0.4

From Table 4.19, it can be seen that the results obtained from the analysis of content pages accessed and time spent indicated that there were no significant differences in average silhouette width when the number of clusters was increased from 3 to 10. Similarly, the results obtained from the analysis of time spent on content indicated no significant differences in average silhouette width as the number of clusters was increasing from 3 to 10. Interestingly, PAM recorded little improvement with the analysis of variables related to the time spent on content. The results obtained from the analysis of only the number of content pages accessed indicated that there was little improvement in silhouette width when the number of clusters was increased from 3 to 10. Overall, there was an improvement in average silhouette when variables related to the number of content pages accessed were analyzed separately from those associated with the time spent. Figure 4.5 shows the silhouette plot for LS variables based on the PAM cluster algorithm.

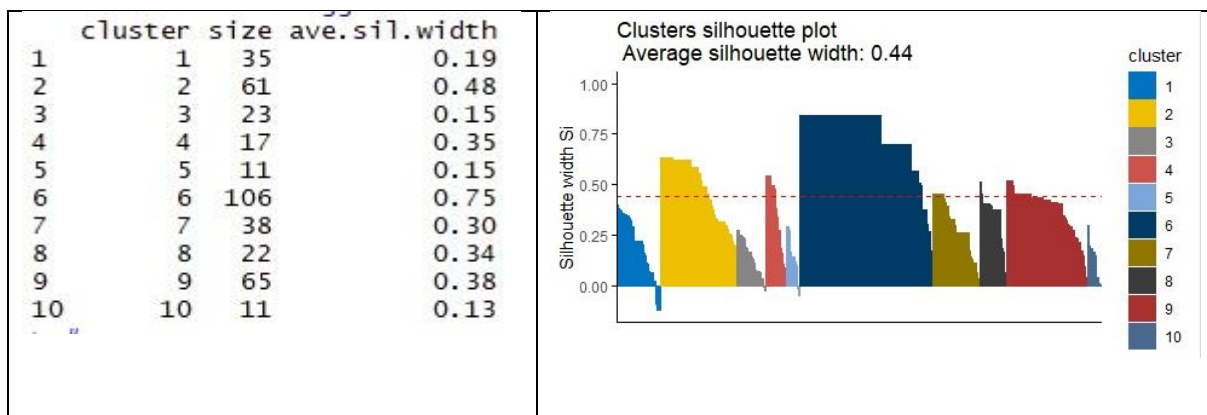


Figure 4.5. Silhouette Plot for Variables Related to LS ( $k=10$ )

From figure 4.5 above, it can be seen that the strongest evidence of the average silhouette width was recorded on cluster 6. Interestingly, cluster 6 contained the maximum number of students. Clusters 5 and 10 had the least number of students but different average silhouette widths. Overall, the average silhouette width was 0.44. This analysis revealed that the compactness, connectedness, and separation of the cluster partitions were reduced when LS variables were used.

### 4.3.3 Results for Hybrid Variables

Table 4.20 shows the first set of generated clusters of variables related to LS and CT. The second set generated clusters for variables related to CT and time spent on contents matching LS. The third set generated clusters for content related to LS and CT.

Table 4.20: Cluster Analysis Results for Variables related to LS and CT

CLUSTERS	% content access, time, and CT variables				Time on content and CT variables				% content access and CT variables			
	k-means	PAM	Diana	HC	k-means	PAM	Diana	HC	k-means	PAM	Diana	HC
3	0.19	0.18	0.19	0.18	0.17	0.16	0.17	0.14	0.22	0.2	0.26	0.23
4	0.17	0.16	0.17	0.19	0.16	0.16	0.16	0.15	0.23	0.2	0.23	0.23
5	0.17	0.15	0.17	0.16	0.17	0.17	0.16	0.14	0.21	0.2	0.23	0.22
6	0.18	0.16	0.17	0.12	0.19	0.16	0.14	0.12	0.21	0.2	0.22	0.21
7	0.18	0.17	0.17	0.1	0.17	0.14	0.13	0.14	0.23	0.2	0.2	0.21
8	0.19	0.16	0.17	0.11	0.18	0.15	0.14	0.15	0.24	0.2	0.2	0.2
9	0.19	0.17	0.17	0.12	0.19	0.14	0.15	0.15	0.24	0.2	0.18	0.19
10	0.17	0.14	0.15	0.13	0.18	0.15	0.13	0.16	0.24	0.2	0.19	0.2

As shown in Table 4.20 above, the results obtained from the analysis of content pages accessed, time spent, attempts on tests, marks scored, and average percentage content coverage indicated that there were no significant differences in average silhouette widths on all algorithms when the number of clusters was increased from 3 to 10. Similarly, the results obtained from the analysis of time spent on content, attempts on tests, marks scored and average percentage content coverage

indicated no significant differences in average silhouette width on all algorithms as the number of clusters was increased from 3 to 10. The results obtained from the analysis of the content pages accessed, attempts on tests, marks scored and average percentage content coverage indicated no significant change in average silhouette width as the clusters increased from 3 to 10. Overall, a significant improvement was observed in the average silhouette when the variables related to content access were analyzed separately. Figure 4.6 shows the silhouette plot for hybrid variables based on the PAM cluster algorithm.

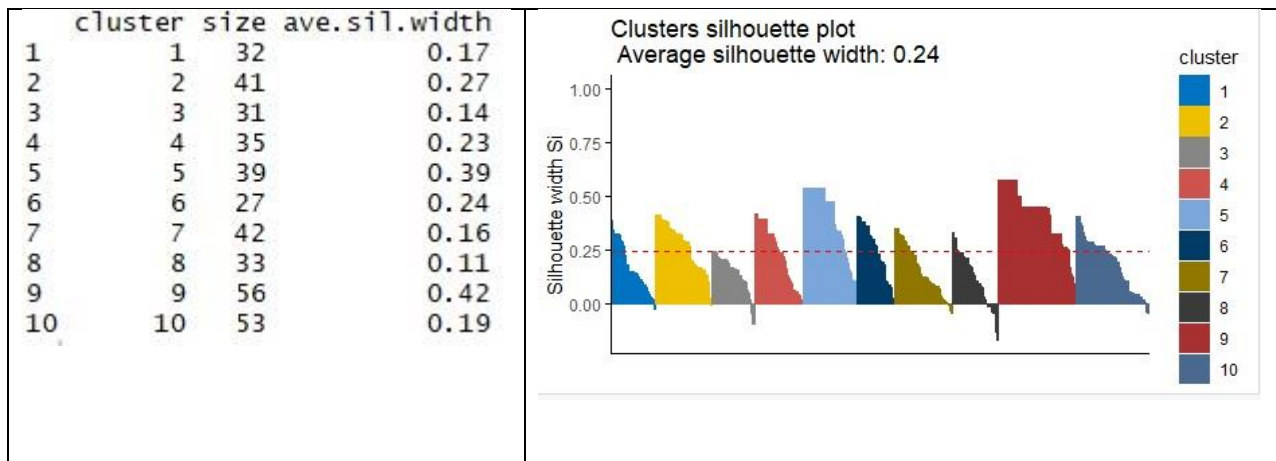


Figure 4.6. Silhouette Plot for Hybrid Variables ( $k=10$ )

From figure 4.6 above, it can be seen that the strongest evidence of the average silhouette was recorded in cluster 9 with 56 students. The same cluster had a maximum number of students. Cluster 6 had the least number of students.

Interestingly, there were also differences in the distribution of students in clusters. Unlike the variables related to CT and LS shown in Tables 4.15 and 4.16 above respectively, the analysis of hybrid variables indicated a uniform distribution of members in each group.

Overall, the average silhouette was 0.24. This analysis revealed that the compactness, connectedness, and separation of the cluster partitions reduced further when both variables related to learning styles and cognitive traits were combined.

Appendix 35 shows the cluster distribution for 20 students. As shown in the cluster table, the rows indicate individual students. The columns are cluster numbers. The value 1 indicates a student belongs to the corresponding cluster number.

#### **4.4 Discussion of Results**

In chapter 2, different theories of LS and CT were introduced and manifestations traits in an LMS were enumerated. In chapter 3, some manifestations of traits were able to be translated into implementation patterns. The identified implementation patterns were coded and implemented in a learning behavior model to analyze student behaviors. This section discusses the evaluation results for the three modeling approaches used in this study.

##### **4.4.1 Discussion of Evaluation Results for the Proposed Model**

The evaluation was done by comparing the results generated by the proposed model and those from the psychometric means. The study used the ILS as a psychometric measure of LS. The performance of online psychometric tests was used to measure CT. The results of each of the four dimensions of LS and CT are discussed.

##### **Evaluation Results for LS**

In evaluating LS estimation by the proposed model, the results from the ILS questionnaire were compared with those from the automatic student modeling approach by CogniLearn.

##### ***Reflective–reflective LS***

Table 4.6 shows the correlations between results predicted by the model and those from the ILS questionnaire concerning active-reflective LS. This study assumed that those who view content with activities and exercises are better active learners. While those who view contents with summaries, self-assessments, and conclusions after reading are better reflective learners. The model proposed in this study estimated active-reflective LS from these patterns. There was a significant correlation of 0.260 ( $p < 0.018$ ) between the results predicted by the model and those gathered from an ILS questionnaire. The positive correlation indicates that those who are likely to visit activities and exercises are better active learners. While those who are likely to visit summaries, self-assessments, and conclusions are better reflective learners. The positive correlation confirms what the current study had assumed. That is, visits to activities and exercises correlate with active LS from an ILS questionnaire. Likewise, visits on summaries and conclusions correlate with reflective LS from an ILS questionnaire. These findings are in agreement with Khan *et al.* (2019) who conducted a similar study and compared the results with those gathered from an ILS questionnaire. The study analyzed the number of self-assessment pages visited as patterns for active LS. The time spent reviewing self-assessment tests was analyzed as patterns for reflective LS. The study reported 0.81 precision on active reflective LS. The study reported a significant positive correlation.

### ***Sensing–Intuitive LS***

The correlations between results predicted by the model and those from the ILS questionnaire concerning sensing–intuitive LS are shown in Table 4.7. There was a significant correlation of 0.595 ( $p < 0.002$ ) between the results. The current study assumed that those who view content with definitions and meanings are better-sensing learners. While those who view content with definitions and meaning are better intuitive learners. The model proposed in this study estimated

sensing-intuitive LS from these patterns. The positive correlation indicates that those who are likely to view topics and sub-topics are better at sensing learners. While those who are likely to view definitions and meanings are better intuitive learners. The positive sign corroborates what the study had anticipated. That is, visiting topics and sub-topics correlate with sensing LS from an ILS questionnaire. Likewise, viewing content pages with definitions and meanings correlate with reflective LS from an ILS questionnaire. These findings are consistent with Khan *et al.* (2019) who developed a similar model and compared the results generated with those gathered from an ILS questionnaire. The study analyzed the number of content pages with examples visited and time spent as patterns for estimating sensing LS. The time spent viewing definitions and meanings was analyzed as patterns for estimating intuitive LS. The study reported 0.781 precision on sensing-intuitive LS. The study reported a significant positive correlation.

### ***Sequential - Global LS***

Table 4.8 shows the correlations between results predicted by the model and those from the ILS questionnaire concerning sequential-global LS. For sequential-global LS, there was a significant correlation of 0.326 ( $p < 0.004$ ) between the results. The positive correlation indicates that those who are likely to view introductions, and overviews are better global learners. While those who are likely to view contents step by step are better sequential learners. The positive sign is an indicator of what the current study had anticipated. That is, introductions and overviews correlate global LS from an ILS questionnaire. Likewise, step-by-step navigation of contents correlates with sequential LS from an ILS questionnaire. The findings on sequential-global LS evaluations are consistent with Khan *et al.* (2019) who conducted a similar study and compared the results with those gathered from an ILS questionnaire. Khan *et al.* analyzed step-by-step content navigations as patterns for estimating sequential LS. The time spent viewing course outlines and



overviews were used as patterns for estimating global LS. Khan *et al* reported 0.78 precision on sequential-global LS. In the current study, the dataset on the content navigation order was used to estimate sequential LS. The number of pages with content outlines and overviews viewed was used to estimate global LS. The study reported a significant positive correlation.

While the evaluation results for the models related to LS, cited in the literature, were measured using precision, the current study uses the kappa statistic method to compare the prototype performance against the psychometric methods.

### **Evaluation Results for CT**

The current study administered the four categories of cognitive tests on a university LMS. Paired associate tests, spatial span tests, abstract reasoning tests adapted, and mental speed tests were used to measure the cognitive behavior of the students. Performance was indicated by the average number of correct answers.

#### ***Associative Learning Ability (ALA)***

In the current study, the proposed model used the dataset on the number of content pages revisited to estimate ALA. Paired associate tests were administered to confirm the associative learning ability of the students. The paired-associate tests tested the ability to remember and recall specific events, paired with the context in which they occurred. The MCQ questions administered required participants to remember what had previously been seen along with the location. The evaluation was by comparing the ALA value approximated by the proposed model with the performances in paired-associative tests. As shown in Table 4.9, the current study reported an insignificant correlation of 0.012 with  $p < 0.640$  between the values estimated by the proposed model and

paired-associate tests. The findings on ALA evaluations are slightly different from Lin, Kinshuk, and Graf (2007) who analyzed versatile content navigations as patterns for estimating ALA. Lin, Kinshuk, and Graf reported a significant correlation of 0.653,  $p < 0.06$  between the value estimated by the proposed model and the results of the computerized psychometric tests. Even though the current study reported an insignificant correlation of 0.012 with  $p < 0.640$ , a marginal positive correlation indicates those with low ALA and who are less likely to revisit learning content are better associative learners. The positive sign confirms what the current study had anticipated. That is, revisiting learning objects correlates with performance in paired-associate tests.

### ***Information Processing Speed (IPS)***

In the current study, the proposed model used the dataset on the number of attempts and grades on online tests to estimate IPS. Mental speed tests were administered to confirm the information processing speed. The questions used were adapted from the psychology today website (Perina *et al.*, 2021). The mental speed tests measured how quickly a student could process information and make decisions based on the available information. For each timed TF question a graphical word/image pair was provided. A student was expected to follow the specific instructions to attempt the questions. Performance was indicated by the number of correct answers. The evaluation took the form of comparing the IPS value approximated by the proposed model with the performances in mental speed tests. From table 4.10 it can be observed that there was a near-significant correlation of 0.455 with  $p < 0.067$  between the value estimated by the proposed model and the mental speed tests. The positive correlation indicates those with high IPS are likely to have few attempts and score high on exercises requiring recall and are better at inductive reasoning. The

positive sign affirms what the current study had anticipated. That is, attempts and marks scored on exercises correlate to performance in mental speed tests.

### ***Inductive Reasoning Ability (IRA)***

In the current study, the proposed model used the dataset on the number of content pages with examples visited to estimate IRA. Abstract reasoning tests were administered to confirm the inductive reasoning ability of the students. The abstract reasoning tests measured the cognitive ability to apply rules to information to arrive at a logical conclusion. The MCQ test required reasoning about the features of several shapes to deduce the one shape that did not fit in with the rest. The evaluation was undertaken by comparing the IRA value approximated by the proposed model with the performances in abstract reasoning tests. As shown in Table 4.11, the findings on IRA evaluations are slightly higher than those of Lin (2007) who analyzed the relationship between visiting pages with examples and the results from a web-based psychometric test with a significant correlation of 0.34,  $p < 0.079$ . The current study reported a significant correlation of 0.224 with  $p < 0.052$  between the value estimated by the proposed model and the abstract reasoning tests. The positive correlation indicates those with high IRA are likely to view examples, revision, and exercises are better at inductive reasoning. The positive sign confirms what the current study had expected. That is, viewing examples, revision, and exercises correlate with performance in abstract reasoning tests.

### ***Working Memory Capacity (WMC)***

In the current study, the proposed model used the dataset on the number of content pages visited linearly to estimate WMC. Spatial span tests were administered to confirm the working memory capacity of the students. The spatial span tests measured the ability to remember the relationships between objects in space. The MCQ questions administered required the participants to remember

the correct sequence boxes throughout the interval. The evaluation took the form of comparing the WMC value approximated by the proposed model with the performances in spatial span tests. The correlation between the WMC value approximated by the proposed model with the performances in spatial span tests at  $-0.087$  with  $p < .561$  is not significant as indicated in Table 4.12 since it shows a negative value. This means a student who has higher WMC should in theory have less reverse navigation (Lin, 2007). The data support the effectiveness of the proposed implementation pattern that forward and backward navigations on an LMS are signs of WMC. That is, forward and backward navigations on an LMS correlate performance in spatial span tests.

#### **4.4.2 Discussion of Evaluation Results for the Neural Network Model**

From the results, when all the LS and CT variables were used, the best prediction results were returned by a network with combinations of 5-3-2 hidden nodes. The precision, recall, f-score, and AUC results for the network with 5-3-2 layer configurations were 0.63, 0.67, 0.64, and 0.65 respectively. When the variables related to LS only were used, the results returned by three-layer configurations range between 0.63 and 0.59 for maximum and minimum respectively. When variables related to CT only were used, the precision, recall, and f-score, the results for a single layer configuration range between 0.67 and 0.43 for maximum and minimum respectively. There were no significant changes recorded above the three-layer configurations.

These results show that the ANN performance is better when the network layers are fewer. These findings contradict the results of Azadeh and Behshtipour (2008) that indicated increasing the number of hidden nodes minimizes the root means square error (RMSE). While the prediction results in Azadeh and Behshtipour's study are measured in terms of RMSE, this study instead used accuracies, precision, recall, and f-score. The prediction results of 0.63, 0.67, 0.64, and 0.65 for

accuracies, precision, recall, and f-score respectively are somehow lower than those reported in the previous studies.

The average performance level was slightly lower compared to those of Renato (2020). Renato's model was based on input variables comprising data collected from an LMS. The assessment attempts and posts viewed on forums were used as indicators for active-reflective LS. The number of contents with facts, definitions, and examples was used as indicators for sensing intuitive LS. Content navigation order and visits on contents with course overviews were used as indicators for sequential global LS. The number of illustrations viewed and visits on contents with texts were used as indicators for visual-verbal LS. Data collected from the ILS questionnaire was used as output for training the model. The model obtained 0.91, 0.81, 0.83, 0.90 from AUC plots for the active-reflective (processing), sensing-intuitive (perception), visual-verbal (input), and sequential-global (understanding) respectively.

Although the current study reported lower prediction accuracy, Renato's results were based only on LS which is a single trait.

One major limitation of the model used in this study is that many input variables were used for the training. The presence of LS and CT variables that have minimal or no significance on the network performance could have affected the results. The presence of outliers not only lowers data quality but also affects the output of a prediction model (Sandbhor and Chaphalkar, 2018). However, it is not possible to eliminate such variables because each is representing a specific LS and CT behavior. Overall, the training results of the model proposed in this study showed that the ANN was able to identify students' cognitive traits and learning styles with high accuracy.

#### **4.4.3 Discussion of Evaluation Results for the Cluster Analysis**

There was strong evidence of the compactness, connectedness, and separation of the cluster partitions detection when the variables related to cognitive traits only were analyzed. The results indicated that when the number of clusters was set to 10, the average silhouette width was 0.85 when variables related to cognitive traits only were analyzed. The average silhouette width was reduced to 0.44 when variables related to learning styles only were analyzed. The average silhouette decreased to 0.24 when all the variables (both LS and CT) were analyzed. Further analysis showed that variables related to content navigations had better results in the cluster partitions than those associated with time spent on the content.

Using a hierarchical approach, Abdelhadi, Ibrahim, and Nurunnabi (2019) built a classifier model using the dataset collected from the ILS questionnaire issued to 44 students to create groups of possible combinations of FSLSM dimensions. The study reported between 0.93 and 0.262 maximum and minimum similarities respectively. The current study evaluated the cluster performance based on the compactness, connectedness, and separation of the cluster partitions.

With regards to LS, the average performance level reported for the hierarchical approach in the current study is 0.2 on variables related to contents accessed and time spent. The average performance is 0.4 on variables related to contents accessed only. The average performance for variables related to CT was between 0.53 and 0.83 at the minimum and maximum respectively. The average performance for variables related to both LS and CT was between 0.2 and 0.4 at the minimum and maximum respectively.

While Abdelhadi, Ibrahim, and Nurunnabi analyzed the dataset gathered from a manual ILS questionnaire, the current study used records of learner interactions with an LMS. Both studies

analyzed features, which are recognized based on FSLSM to classify LS as active–reflective, sensing–intuitive, sequential–global, and visual-verbal.

Using a k-means approach, Fatahi, Moradi, and Farmad (2015) built a classifier model using the dataset collected from browsing records of 198 students to create groups of possible combinations of MBTI LS. The browsing records analyzed were response time on tests, folder visits, homework solution visits; forum visits; messages posted; time of the first view of the file since uploading. The study reported f-score results of 0.64, 0.60, 0.36.0, 0.36, 0.74, 0.54, 0.63, 0.36.0 for an introvert, extroverts, sensing, intuition, thinking, feeling, perceiving, and judging respectively. While the prediction results in Fatahi, Moradi, and Farmad’s study are measured in terms of f-score, the current study used the compactness, connectedness, and separation of the cluster partitions instead.

The current study analyzed features, which are recognized based on FSLSM to classify LS as active–reflective, sensing–intuitive, sequential–global, and visual-verbal. With regards to LS, the average performance level reported for the k-means approach in the current study is 0.2 on variables related to contents accessed and time spent. The average performance is 0.4 on variables related to contents accessed only.

The average performance for variables related to CT was between 0.5 and 0.81 at a minimum and maximum respectively. The average performance for variables related to both LS and CT was between 0.17 and 0.24 at a minimum and maximum respectively. These results seemed to be slightly lower compared with other learner behavior models in the existing literature.

These results, in general, suggest that the average silhouette width reduces when the number of variables increases. A possible explanation for this might be that the number of variables used in the analysis increased data collinearity (Lindner, Puck, and Verbeke, 2020).

These findings could be attributed to the higher levels of correlations among variables which can cause problems in segmentation analysis. The dataset containing variables for cognitive traits when analyzed separately generated the highest silhouette width. The data collinearity decreased when variables were few. The fewer the variables, the less effect it had on segmentation analysis. For the dataset with both learning style and cognitive traits variables, there was no significant change in silhouette width when the number of clusters was increased or decreased. This is could be because the variables were over ten and this increased data collinearity.

It is important to note that the two studies cited are based on single learning behavior, while the current study is based on both single and hybrid traits of LS and CT.

The overall results of the evaluation analysis showed a positive silhouette co-efficient for all clusters. The positive silhouette co-efficient values generally indicated that all the students had been assigned to the right clusters. These results confirm the existing evidence that learner navigation behavior is a predictor variable for a personality model such as LS or CT. The results confirm that learner navigation patterns in an LMS are indicators of personality models described in the literature.

#### **4.5 Comparison of the results regarding the three approaches**

Table 4.21 report the results of LS and CT prediction using the Cognilearn model. The results given for LS indicate fair agreements for active-reflective and sequential–global learning styles. The results for sensing-intuitive are moderate. The results given for CT indicate slight agreement



for associative learning ability. The results for information processing speed and inductive reasoning ability are moderate and fair respectively. The results for working memory capacity report no agreement with those predicted by the Cognilearn model.

*Table 4.21. Evaluation Results for the Cognilearn Model*

	Behavior	Kappa results	classification
Variables related to LS	Active-reflective	0.260 (p < 0.018)	fair
	Sensing - intuitive	0.595 (p < 0.002)	moderate
	Sequential – global	0.326 (p < 0.004)	fair
Variables related to CT	Associative learning	0.012 (p < 0.640)	slight
	Information processing speed	0.455 (p < 0.067)	moderate
	Working memory capacity	-0.087 (p < 0.561)	no agreement
	Inductive reasoning ability	0.224 (p < 0.052)	fair

Table 4.22 report the results of LS and CT prediction using the ANN model. The results given for variables related to LS indicate 0.67, 0.61, 0.59, and 0.63 maximum precision, recall, f-score, and AUC respectively. The results given for variables related to CT indicate 0.60, 0.67, 0.63, and 0.62 maximum precision, recall, f-score, and AUC respectively. These results can be interpreted as moderate prediction accuracy.

*Table 4.22. Evaluation Results for the ANN Model*

Behavior	maximum precision	maximum recall	maximum f-score	maximum AUC
LS variables	0.67	0.61	0.59	0.63
CT variables	0.60	0.67	0.63	0.62

Table 4.23 report the results of LS and CT prediction using the cluster model. The results given indicate 0.44 and 0.88 average silhouette width for variables related to LS and CT respectively.

*Table 4.23. Evaluation Results for the Cluster Model*

Behavior	Average silhouette Width
LS variables	0.44
CT variables	0.88

The evidence from the study suggests that the cluster model yields better results than the Cognilearn model and ANN model regarding the prediction of CT. ANN model yields better results than the Cognilearn model and cluster model regarding the prediction of LS. Both the

cluster model and ANN model yield better results than the Cognilearn model regarding the prediction of LS and CT.

The results are inconsistent with the findings of Khan *et al.* (2019) in which the results of the machine learning approach using a Bayesian network for inferring learning styles and affective states from an LMS were compared with the rule-based method. The study reported that the rule-based approach yielded better results than the machine-learning approach. It seems possible that these results are inconsistent because the study conducted by Khan *et al.* (2019) analyzed variables related to learning styles and affective states. The current study analyzed the variables related to learning styles and cognitive traits. The different psychological variables applied may explain the inconsistency in the results. Another possible explanation is the different evaluation parameters used. For machine learning evaluation, Khan *et al.* (2019) made comparisons between the learning behavior predicted by the Bayesian network and the ILS questionnaire. Instead of using data collected from the ILS questionnaire to train the machine learning model, the current study used that dataset collected from the psychology module as input and the information generated from the sociology module as output for training the ANN model. The different output datasets used to train the machine learning models may explain the inconsistency in the results.

#### **4.6 Summary**

Evaluation results for the proposed model indicated moderate agreement between the LS results predicted by the model against those gathered through the ILS questionnaire. Except for WMC which indicated a disagreement, evaluation results for the proposed model showed moderate agreement between the CT results predicted by the model against those gathered through the computerized psychometric tests administered online.

Evaluation results for the ANN reported that the best accuracy, precision, recall, f-score, and kappa results for the network were 0.63, 0.67, 0.64, 0.65, and 0.26. The kappa results are indications that the collected data moderately represents the LS and CT variables investigated

Overall results of cluster evaluation analysis showed a positive silhouette coefficient for all clusters. The positive silhouette coefficient values generally indicated that all the students had been assigned to the right clusters.

These results confirm the existing evidence that learner navigation behavior is a predictor variable for a personality model such as LS or CT. This amplifies the relationship between learner navigation patterns in an LMS and personality models provided in the existing literature.

## **CHAPTER 5: CONCLUSION AND RECOMMENDATIONS**

The study set out to investigate and propose an automatic method for identifying LS and CT with high accuracy in an LMS guided by the following research questions:

- 1** What are the available methods applicable to the automatic detection of learner behavior in an LMS?
- 2** How can the feature sets for the automatic detection of learner behavior be extracted from LMS logs?
- 3** How can a computing model be used to identify a hybrid learner behavior in an LMS?
- 4** How can the above computing model be evaluated?

The first section of this chapter discusses the summary of the findings in the context of the appropriate LS and CT identified from the existing literature. It also outlines the attributes used for automatic modeling methods and the evaluation techniques applied. The second section discusses the significance of the findings from this study. The third section discusses the limitations of this. Finally, the study recommends areas for further research.

### **5.1 Summary of the Findings**

The findings are discussed under the respective research questions.

#### **5.1.1 The available methods that are applicable for the automatic detection of learner behavior**

The current study reviewed related works as presented in Chapter 2. It established that learner behavior could be identified in an LMS by applying traditional manual psychometric tools based on various psychological theories. For example, the Index of Learning Style (ILS) is a questionnaire for collecting data based on FLSM (Felder, 1988). Honey and Mumford's Learning Styles Questionnaire (LSQ) for collecting data based on Honey and Mumford LSM (Honey and

Mumford, 1982). The Operation Span (OSPAN) test for collecting data on memory capacity (Turner and Engle, 1989). However, these manual tools are prone to errors and require extra effort from a learner to obtain information regarding their learning preferences (Kay and Kummerfeld, 2012).

Automatic methods that analyze records of learner interactions with an LMS to infer their behavior have been proposed to overcome the problems associated with the use of manual questionnaires and psychometric tests. To identify learner characteristics in an LMS, the human observable indicators are extracted for analysis to create learner profiles and estimate the hints for a particular psychological behavior. The previous studies reviewed that apply automated methods include the examples:

- Using Bayesian network analysis, Conditional Probability Tables (CPT) for each learning style attribute (Rajper *et al.* 2016).
- Using the intelligent agent method, a model can initialize, learn, match, and recommend the learning styles of an individual (Hasibuan *et al.* 2016).
- Using a decision tree algorithm, a classification technique that can predict LS by analyzing log data extracted from an LMS (Liyanage, Gunawardena, and Hirakawa, 2016).
- Using neural network analysis, records of students' interactions with an LMS can be fed into the network as input and the specific learning attributes can be used as an output to be predicted by the network Aditya and Babita (2018).
- Using the simple rules, the actions of learners on an LMS can be extracted to get hints about their learning preferences, and then apply a simple rule-based method to estimate the behavior from the number of matching hints (Khan *et al.*, 2019).

Some automated methods only classify students according to a single psychological behavior e.g. learning preferences (Abdelhadi, Ibrahim, and Nurunnabi, 2019). The problem of a single behavior is either an incomplete or partial student modeling process and such as unreliable information. Some automated methods classify students according to multiple psychological behaviors e.g. learning preferences and affective states (Khan *et al.*, 2019), and anxiety, personality, and learning preferences (Aditya and Babita, 2018).

However, the techniques of these automated methods for identifying feature sets from LMS logs and relating them to different theories that describe learning behaviors are still inconsistent and inconclusive.

The current study introduced automated hybrid learner behavior models integrating characteristics related to learning styles and cognitive trait models. The hybrid models are designed according to the descriptions of the dimensions that appear in the learning style and cognitive theories.

### **5.1.2 How Feature Sets for Automatic Detection of Learner Behavior can be Extracted from LMS logs**

The study identified indicators for LS preferences from LMS logs based on FSLSM and grouped them as:

- **active:** visits and time spent on content with activities
- **reflective:** visits and time spent on content with summaries and conclusions
- **sensing:** visits and time spent on topics and subtopics
- **intuitive:** visits and time spent on contents with definitions
- **visual:** visits and time spent on content with illustrations
- **verbal:** visits and time spent on content with text
- **sequential:** pages navigated without skipping

- **global:** visits and time spent on content with outlines, introductions, and overviews.

The study identified indicators for CT preferences from LMS logs based on CTM and grouped them as:

- information processing speed: attempts and marks obtained on exercises
- inductive reasoning ability: visits and time on content with examples
- working memory capacity: the percentage of content accesses out of the total
- associative learning ability: the number of revisited contents pages.

In the discussion under section 3.2.1, SQL search queries with keywords related to the indicators of LS and CT were used to extract patterns from the browsing records in the LMS database. This was as follows.

- A query containing the search keyword ‘activity’ was used to extract records on the number of learning objects viewed and time spent related to active LS. Similarly, records related to reflective LS were extracted using a query containing the search keywords ‘summary’, ‘self-assessments’, and ‘revision’.
- A query containing the search keywords for identifying sub-topics with facts and details was used to extract patterns on the number of learning objects viewed and time spent related to sensing LS. Similarly, records related to intuitive LS were extracted using a query containing the search keywords ‘definitions’, ‘self-assessments’, and ‘revision’.
- A search query with a keyword to identify ‘figure’, ‘illustration’, ‘picture’, etc. accessed was used to extract patterns on the number of learning objects viewed and time spent related to visual and verbal LS.
- A search query with the keyword identifying ‘introductions’ and ‘overviews’ accessed by a student was used to extract patterns on the number of learning objects viewed and time

spent related to global LS. Likewise, records related to sequential LS were extracted using a search query with the keyword ‘content navigation order’.

- A search query for counting the same learning objects revisited on different dates and times by each student was used to generate patterns for ALA.
- A search query with the keywords ‘example’ or ‘case studies’ was to extract patterns on the number of learning objects viewed and time spent related to IRA.
- A search query for counting the number of learning objects accessed once by a student considering percentage content coverage was used to extract patterns related to WMC.
- A search query for counting the number of attempts and computing average marks scored on exercises by each student was used to extract patterns related to IPS.

From these SQL search query activities with keywords related to the indicators of LS and CT, navigation records were extracted and analyzed to identify patterns related to LS and CT. This addressed the second research question on how the feature sets for the automatic detection of learner behavior could be extracted from LMS logs.

### **5.1.3 How a Computing Model can be used to Identify Learner Behavior in an LMS**

The current study addressed this question by designing and developing a rule-based model CogniLearn with extraction engines for fetching relevant patterns associated with LS and CT. From this, an estimation function for computing LS and CT was implemented to calculate the behavior of each student and map it to a 3-item scale: 0.1 - 0.3 low, 0.4-0.6 Moderate, 0.7 – 1.0 high, and 0.0 – no preference. The model was enhanced to match and display the learning contents relevant to the individual behavior estimated for each student. CogniLearn provides a unified automated learner behavior model integrating characteristics from FSLSM and CTM. CogniLearn model estimates learning preferences based on direct observation of learner behavior patterns



which addresses the problem of psychometric flaws of traditional manual instruments such as the ILS questionnaire.

To compare the performance of the model in a different environment, the dataset collected from LMS logs was analyzed using machine learning methods.

The current study developed an artificial neural network model. To implement an artificial neural network model, the records of learner interactions with contents on psychology were used as input for training the network. The records of learner interactions with contents on sociology were processed using if-else rules to get the output values for training the network. The dataset was entered into the network and trained to classify learners according to LS and CT. The prediction accuracy was evaluated using accuracy, precision, recall, f-score, and area under the curve (AUC).

Further, developed a cluster model. To implement a cluster-based model, the dataset generated from the records of learners' interactions with an LMS was used to analyze the records of the students. From this task, a cluster-based model was designed and developed to profile learners with similar LS and CT. The k-means, hierarchical, PAM, and Diana algorithms were used to create clusters. The silhouette analysis was used to assess the quality of the clusters created. The study implemented a nearest-neighbor model to match the learner preferences to the existing groups in the system. This is following the proposal that the filtering algorithm can be used to recommend learning content by considering a group with similar preferences (Romero et al., 2007). The current study assumed that students belonging to the same clusters share similar characteristics.

#### **5.1.4 How the Performance of the Model can be Evaluated**

Related studies proposed that the performance of an adaptive learning system can be evaluated by comparing it with a non-adaptive version of the system (Weibelzahl, 2005; Paramythis and Weibelzahl, 2005). This can be achieved by matching the prediction results with data from the LS questionnaire- through an evaluation of students (Lamia and Yamina, 2018). Drissi and Abdelkrim (2017) indicated that adapting teaching strategies based on data from the LS questionnaire can be experimentally evaluated with students to test match/mismatch.

To evaluate the performance of the CogniLearn, the current study applied the interrater reliability test to measure the agreement between the results gathered using the manual methods and the model. The interrater reliability test uses kappa statistics as the metric method for assessing the agreement between two raters (Cohen, 1960).

The kappa statistic method was applied to determine consistency among the two raters. The raters used in the current study were manual psychometric methods against the CogniLearn model. The results collected through Index of Learning Styles Questionnaires (ILS) and Online Cognitive Multiple Choice Questions administered to students were compared against those generated by the CogniLearn model. The evaluation results are discussed in section 4.1.

In evaluating the performance of the developed artificial neural network, the dataset was analyzed in three groups:

- All variables related to LS and CT
- Variables related to LS
- Variables related to CT

The matrices commonly applied to evaluate the prediction accuracy of a machine learning model were applied in the current study. The methods are accuracy, precision, recall, f-score, and area under the curve (AUC) (Zheng, 2015). Precision was applied to measure the ratio of correctly predicted positive observations to the total predicted positive observations. The recall was used to measure the ratio of correctly predicted positive observations to all observations in the actual class. F1-score was obtained by calculating the weighted average of Precision and Recall. The evaluation results are discussed in section 4.2.

In evaluating the performance of the developed cluster model, as presented in section 3.5.2 the dataset was analyzed in three groups:

- All variables related to LS and CT
- Variables related to LS
- Variables related to CT

The silhouette coefficient method was used to measure the qualities of the clusters created. The technique measures the compactness, connectedness, and separation of the cluster partitions (Brock *et al.*, 2008). Connectedness is the ability to place group members in the same cluster as their nearest neighbors in the data space. Compactness is the assessment of the intra-cluster variance. Separation is the assessment of the degree of separation between clusters based on the distance between their centroids.

The first set of evaluations examined the impact of variables related to CT on cluster quality. The second set examined the impact of variables related to LS on cluster quality. The last set examined the impact of variables related to LS and CT on cluster quality. The dataset was analyzed by creating clusters between three and ten. The average silhouette was recorded for each number of clusters at each stage. The evaluation results are discussed in section 4.3.

## **5.2 Research Contributions**

This study makes several significant contributions to the field of learner modeling in adaptive educational systems. This section highlights the contributions of this study.

### **5.2.1 Practical Contribution**

The current study extends the work of Khan *et al.* (2019) who developed a hybrid adaptive model for the automatic identification of learning styles and affective states in an LMS. From this approach, a tool 'Affective States and Learning Styles Identification and Measurement' (AsLim) can be used by teachers to identify the learning styles and affective states of their learners was implemented. The current study developed a hybrid model of *Cognilearn*. The model applies a rule-based method to calculate learning styles and cognitive traits from records of learners' interactions with an LMS. The model uses indicators of learner preferences based on FSLSM and CTM to infer learning preferences. The model filters and orders learning contents according to estimated LS and CT preferences.

### **5.2.2 Methodological Contribution**

The current study extends the research in the design and development of machine learning models based on the foundations of the work conducted by Aditya and Babita (2018), and Sweta and Lal (2016). These studies apply the records of learner interactions with an LMS as training input for the artificial neural network to predict learning preferences. According to Rojas (1996), an artificial neural network is a set of interconnected nodes that receive inputs, then process them to generate output. The incoming signals are converted into a single outgoing and then broadcasted to other nodes. The activation of an ANN neuron is determined when the inputs are multiplied by weights and then computed by a mathematical function.

The study aimed at reducing psychometric flaws associated with psychometric tools such as ILS. Data collection questionnaires like ILS are prone to mistakes and tedious to administer. The current study used the records of learner interactions with the psychology module as input for training the network. The data collected from the sociology module were encoded to produce the output for training the network to classify learners according to LS and CT. This would reduce the psychometric flaws associated with manual psychometric tools e.g. ILS.

The study also extends the research in the design and development of machine learning models based on the work of Abdelhadi, Ibrahim, and Nurunnabi (2019), and Poellhuber and Roy (2019). The studies used cluster algorithms to profile students with similar navigation behavior on LMS. According to Han, Kamber, and Pei (2012), clustering is an unsupervised machine-learning technique used to group objects which are closer to one another in a multidimensional feature space usually to uncover some inherent structure that the data possesses.

Abdelhadi, Ibrahim, and Nurunnabi(2019) used an LS questionnaire to get students' LS of students and then used a clustering algorithm to put them in similar groups. Poellhuber and Roy (2019) used a clustering algorithm to group students according to their browsing habits, self-assessments, serious reading, active independence, and active socialization.

The current study undertakes to design and develop a hybrid model that contains multiple LS and CT to improve learner characterization. The study applied four cluster algorithms to analyze the dataset. These algorithms are k-means, hierarchical, PAM, and Diana. The dataset was analyzed in three groups namely variables related to LS, CT, and both. Finally, the silhouette analysis was used to assess the computational efficiency, accuracy, and quality of clusters created. The results indicated that all students were placed in the correct cluster.

### **5.2.3 Theoretical Contribution**

The study extends an approach that is founded on information processing theory by applying computer technology to understand how the mind of a learner processes information. According to Vinney (2020), information processing theory is the foundation of mental psychology applying computer technology to illustrate the processing mechanism of the mind. According to Slate and Charlesworth (1989), information processing theories explain the perception and recall of information by the mind.

The current study extends research work in applying information processing theories to illustrate the processing mechanism of the mind when a student is browsing contents in an LMS. The study identifies indicators of information processing related to LS and CT from the records of learner interactions with an LMS.

Identifying indicators for inferring LS in an LMS extends the studies conducted by researchers e.g. Liyanage, Gunawardena, and Hirakawa (2014); Pham and Florea, (2013). Both studies identified data on the number of visits on exercise and assessment as indicators of active LS. The researchers identified visits and time spent on content with summaries as hints for a reflective LS. This study applied a rule-based method to estimate LS from matching hints. An evaluation was done by comparing the results against those gathered through the ILS questionnaire with a prediction accuracy of 0.634 and 0.72 precision on active-reflective LS.

Identifying indicators for inferring CT in an LMS extends the studies conducted by researchers e.g. Lin, Kinshuk, and Graf (2007). The study identified versatile content navigations as indicators of ALA and then applied a rule-based method to estimate CT from matching hints. An evaluation was done by comparing the results against those gathered through psychometric tests with a significant correlation of 0.653,  $p < 0.06$ .

The current study used a dataset on the number of activities and exercise pages viewed to predict active LS. The number of summaries, self-assessments, and conclusions pages viewed were used to estimate reflective LS. The current study reported a significant positive correlation. The study applied a rule-based method to estimate LS from matching hints based on FSLSM and CTM to improve learner characterization in an LMS.

#### **5.2.4 Contribution to the Field of Information Systems**

The current study has presented a method for analyzing data on learning activities, identifying behavior patterns, and providing information that can be used by an instructor to improve learning (Cambuzzi, Rigo, and Barbosa, 2015). This has been accomplished by embedding the information processing theory of the mind to accommodate students with diversified identities (Van Wingerden, 2021).

CogniLearn model presented in this study supports pedagogical strategies by displaying the learning behavior of a student and recommending appropriate learning materials. Such a platform supports pedagogical strategies by offering recommendations visualization interfaces (Waheed *et al.*, 2020).

Analyzing educational data to generate knowledge that results in the discovery of students' behavior can assist instructors to understand a learner better and apply proper interventions in an online learning environment (Waheed *et al.*, 2020).

Thus the study extends research in the development of learning patterns for identifying learning behavior from the utilization of an LMS. Such information helps an instructor to have a better conception of learning and to improve teaching (Gaševića, Kovanović, and Joksimović, 2017; Clow, 2013). Thus an instructor can better understand the learning behavior of a student, provide appropriate interventions, and improve learning outcomes (Andrade, Rigo, and Barbosa, 2021).

### **5.3 Limitations of the Current Study**

The study was limited by the fact that only preclinical medicine students actively used the LMS to access learning materials. Even though the model can estimate learner behavior and provide appropriate learning content, the study only evaluated the prediction accuracy against the psychometric methods. The current study did not evaluate the impact of the adaptive learning content provided by the model on learning outcomes.

### **5.4 Recommendations for Further Research Work**

Further research regarding learner modeling in learning management systems with students from other fields would be a worthwhile consideration.

Additional studies using the same method to identify other cognitive traits should be carried out. More research is also required to conduct additional evaluation experiments in further evaluation of the model.

Even though the proposed prototype is an adaptive learner model that automatically allocates learning resources that are relevant to the identified LS and CT, further evaluation with students in a classroom environment is required.



## REFERENCES

- Ababneh, S. ' (2015) 'Learning Styles and Preferences of Jordanian EFL Graduate Students', *Journal of Education and Practice*, 6(15), pp. 31–37. Available at: [www.iiste.org](http://www.iiste.org).
- Abbot, B. (2002) *Human memory.*, Fort Wayne: Indiana University-Purdue University at Fort Wayne, Psychology Department. Available at: <http://users.ipfw.edu/abbot/120/LongTermMemory.html>.
- Abdelhadi, A., Ibrahim, Y. and Nurunnabi, M. (2019) 'Investigating engineering student learning style trends by using multivariate statistical analysis', *Education Sciences*, 9(1). doi: 10.3390/educsci9010058.
- Adadi, A. and Berrada, M. (2018) 'Peeking inside the black-box: a survey on explainable artificial intelligence', in *IEEE Access*.
- Adebayo, E. I. and Wokocha, C. M. (2011) 'A guide to establishing management information system in tertiary institutions in nigeria', in *Proceedings of the 2011 International Conference on Teaching, Learning and Change International Association for Teaching and Learning (IATEL)*, pp. 54–62.
- Aditya, K. and Babita, P. (2018) 'SVM and PCA Based Learning Feature Classification Approaches for E-Learning System', *International Journal of Web-Based Learning and Teaching Technologies*, 13(2), pp. 32–45.
- Agustiono, W. (2018) 'Applying Evolutionary Prototyping In Developing LMIS : A Spatial Web- Based System For Land Management Applying Evolutionary Prototyping In Developing LMIS : A Spatial Web-Based System For Land Management', in *The 2nd International Joint Conference on Science and Technology (IJCST) 2017*.
- Ahmad, Z. and Shahzadi, E. (2018) 'Prediction of Students ' Academic Performance using Artificial Neural Network', 40(3), pp. 157–164.
- Al-Omari, M., Carter, J. and Chiclana, F. (2016) 'A hybrid approach for supporting adaptivity in e-learning environments', *International Journal of Information and Learning Technology*, 33(5), pp. 333–348. doi: 10.1108/IJILT-04-2016-0014.
- Allen, E. I. and Seaman, J. (2017) *Digital Compass Learning: Distance Education Enrollment Report 2017*.
- Andersen, H. V. and Sorensen, E. K. (2017) 'Enhancing Understanding , Flow and Self - Efficacy in Learners with developmental and tention Difficulties Through ICT - Based Interventions', 20(1), pp. 153–174.
- Anderson, J. (1985) *Cognitive psvcbology and its applicatlons*. 2nd edn. New York: W.H. Freeman and Company.
- Andrade, T. L. de, Rigo, S. J. and Barbosa, J. L. V. (2021) 'Active Methodology, Educational

Data Mining and Learning Analytics: A Systematic Mapping Study’, *Informatics in Education*, 20(2), pp. 171–204.

Anjum, R. *et al.* (2019) ‘A meta-model to automatically generate evolutionary prototypes from software requirements’, in *Proceedings of the 7 th International Conference on Computer and Communications Management*, pp. 131–136.

Ariffin, A. H. *et al.* (2020) ‘The Development of an e-Workload Distribution System: A Focus on the Fair Distribution of Teaching Workloads of Lecturers’, *Asian Journal of University Education*, 16(4), pp. 31–43. doi: 10.24191/ajue.v16i4.11955.

Assis, L. *et al.* (2022) ‘Relationship Between Learning Styles and Learning Objects: A Systematic Literature Review’, *International Journal of Distance Education Technologies*, 20(1), pp. 1–18. doi: 10.4018/IJDET.296698.

Atkinson, R. and Shiffrin, R. (1968) ‘Human memory. A proposed system and its control processes.’, in Spence, K. and Spence, J. (eds) *The psychology of learning and motivation*. Princeton: Van Nostrand.

*aTutor* (2002). Available at: [atutor.github.io](http://atutor.github.io) (Accessed: 21 December 2020).

Azadeh, A. and Behshtipour, B. (2008) ‘The effect of neural network parameters on the performance of neural network forecasting’, in *IEEE International Conference on Industrial Informatics (INDIN)*. IEEE, pp. 1498–1505. doi: 10.1109/INDIN.2008.4618341.

Baneres, D. *et al.* (2021) ‘A predictive analytics infrastructure to support a trustworthy early warning system.’, *Applied Sciences*. doi: [https:// doi.org/ 10.3390/app11135781](https://doi.org/10.3390/app11135781).

Baumgartner, P., Häfele, H. and Maier-Häfele, K. (2002) *E-Learning Praxishandbuch - Auswahl Von Lernplattformen*. Studien Verlag.

Begum, M. R. and David, K. (2017) ‘Discovering Student Learning Style using Min Max Cascade Neural Network’, *Indian Journal of Science and Technology*, 10(25), pp. 1–15. doi: 10.17485/ijst/2017/v10i25/110081.

Bernard, J. *et al.* (2015) ‘Using artificial neural networks to identify learning styles’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, pp. 541–544. doi: 10.1007/978-3-319-19773-9\_57.

Betz, M. K. (2006) ‘Solo and Social Learning in Online Courses: Implications for Information Processing Theory’, *International Journal of Instructional Technology and Distance Learning*, 3(2).

Biehler, R. and Snowman, J. (1986) *Psychology applied to teaching*. 5th edn. Boston: Houghton Mifflin Company.

Biggs, J. B. and Collis, K. F. (1982) ‘Evaluating the quality of learning: the SOLO taxonomy (structure of the observed learning outcome).’, in *Educational psychology series*. New York: Academic Press.

- Bishop, C. M. (2006) *Pattern Recognition and Machine Learning*. Edited by M. Jordan, J. Kleinberg, and B. Schölkopf. Springer - New York.
- Blakemore, T., McCray, P. and Coker, C. (1984) *A Guide to Learning Style Assessment. Research and Training Center Research Report*.
- Bojte, D. (2019) 'Mastering the Skills Gap at Systemic Level – The Education Management Information System: A Key Element for an Effective Digital Transformation in Education', *Review of International Comparative Management.*, 20(2), pp. 131–143.
- Borkar, P. (2022) *Role of management information system (MIS) in education sector*. Available at: <https://www.iitms.co.in/blog/role-of-management-information-system-in-education.html>].
- Boticki, I., Akçapınar, G. and Ogata, H. (2019) 'E-Book User Modelling through Learning Analytics: The Case of Learner Engagement and Reading Styles', *Interactive Learning Environments*, 27(5–6), pp. 754–765.
- Bouchrika, I. (2022a) *Learning Management Systems for Education: Features, Benefits, and Challenges*. Available at: <https://research.com/software/learning-management-systems-for-education>.
- Bouchrika, I. (2022b) *What is Information Processing Theory? Stages, Models & Limitations*. Available at: <https://research.com/education/what-is-information-processing-theory>.
- Bradford, A. (2017) *Deductive Reasoning vs. Inductive Reasoning, Live Science Newsletter*. Available at: <https://www.livescience.com/21569-deduction-vs-induction.html> (Accessed: 16 October 2021).
- Brenkel, M. *et al.* (2017) 'Assessing capacity in the elderly: comparing the MoCA with a novel computerized battery of executive function.', *Dementia and Geriatric Cognitive Disorders Extra*, 7, pp. 249–256.
- Brock, G. *et al.* (2008) 'cValid: An R Package for Cluster Validation', *Journal of Statistical Software*, 25(4). doi: 10.1016/0038-1098(77)91248-0.
- Brown, A. (1975) 'The development of memory: Knowing, knowing about knowing, and knowing how to know', *Advances in child development.*, 10.
- Brusilovsky, P. and Millán, E. (2005) 'User models for adaptive hypermedia and adaptive educational systems', in Roll, I., Baker, R. S., Aleven, V., McLaren, B. M., & Koedinger, K. R. (ed.) *User modeling*. Berlin Heidelberg: Springer, pp. 367–376. Available at: [http://scholar.google.com.ezp.lib.unimelb.edu.au/scholar?hl=en&lr=&q=info:xGF\\_pSi0ug4J:scholar.google.com/&output=search](http://scholar.google.com.ezp.lib.unimelb.edu.au/scholar?hl=en&lr=&q=info:xGF_pSi0ug4J:scholar.google.com/&output=search).
- Cambridge Brain Sciences* (1998). Available at: [www.cambridgebrainsciences.com](http://www.cambridgebrainsciences.com) (Accessed: 21 December 2020).
- Cambruzzi, W., Rigo, S. J. and Barbosa, L. V. J. (2015) 'Dropout Prediction and Reduction in Distance Education Courses with the Learning Analytics Multitrail Approach', *Journal of Universal Computer Science*, 21(1), pp. 23–47.

- Cassidy, S. (2004) 'Learning styles: An overview of theories, models, and measures', *Educational Psychology*, 24(4), pp. 419–444. doi: 10.1080/0144341042000228834.
- Chang, Y. H. *et al.* (2016) 'Yet another adaptive learning management system based on Felder and Silverman's Learning Styles and Mashup', *Eurasia Journal of Mathematics, Science and Technology Education*, 12(5), pp. 1273–1285. doi: 10.12973/eurasia.2016.1512a.
- Chango, W. *et al.* (2021) 'Improving prediction of students' performance in intelligent tutoring systems using attribute selection and ensembles of different multimodal data sources', *Journal of Computing in Higher Education*. Springer US, (0123456789). doi: 10.1007/s12528-021-09298-8.
- Chen, F. and Cui, Y. (2020) 'Utilizing student time series behaviour in learning management systems for early prediction of course performance', *Journal of Learning Analytics*, 7(2), pp. 1–17. doi: 10.18608/JLA.2020.72.1.
- Chulkov, D. V. (2017) 'On the role of switching costs and decision reversibility in information technology adoption and investment', *Journal of Information Systems and Technology Management*, 14(3), pp. 309–321.
- Clark, R. C. and Mayer, R. . (2011) *E-learning and the science of instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. San Francisco, CA: Pfeiffer.
- Claroline (2017). Available at: <https://www.claroline.net> (Accessed: 21 December 2020).
- Clow, D. (2013) 'An overview of learning analytics', *Teaching in Higher Education*, 18(6), pp. 683–695.
- Cohen, J. (1960) 'A Coefficient of Agreement for Nominal Scales', *Educational and Psychological Measurement*, 20(1), p. 37.
- Correa, V. P., Monsalve, P. J. and Tabares, B. M. (2020) 'A systematic mapping review of context-aware analysis and its approach to mobile learning and ubiquitous learning processes.', *Computer Science Review*, 39.
- Craik, F. I. M. and Lockhart, R. S. (1972) 'Levels of processing: A framework for memory research', *Journal of Verbal Learning and Verbal Behavior*, 11(6), p. 671.
- Creative Commons (2013). Available at: [creativecommons.org](http://creativecommons.org).
- Deary, I. *et al.* (2004) 'The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947', *Journal of Personality and Social Psychology*, 86(1), pp. 130–147.
- Dougiamas, M. (2022) *Moodle - Philosophy*. Available at: <http://docs.moodle.org/en/Philosophy>.
- Dunn, R. (1990) *Rita Dunn Answers Questions on Learning Styles.*, *Educational Leadership*.
- Dunn, R. and Burke, K. (2006) *Learning Style: The Clue To You, LSCY: Research and Implementation Manual*. Available at: 47 [www.cluetoyou.com](http://www.cluetoyou.com).

Dvorak, T. and Jia, M. (2016) ‘Do the Timeliness, Regularity, and Intensity of Online Work Habits Predict Academic Performance?’, *Journal of Learning Analytics*, 3(3), pp. 318–330. doi: 10.18608/jla.2016.33.15.

El-Bishouty, M. M. *et al.* (2015) ‘Analyzing Learner Characteristics and Courses Based on Cognitive Abilities, Learning Styles, and Context’, in Chang, M. and Li, Y. (eds) *Smart Learning Environments. Lecture Notes in Educational Technology*. Springer, Berlin, Heidelberg, pp. 3–25. doi: 10.1007/978-3-662-44447-4\_1.

Eliezer, J. *et al.* (2022) ‘Cypriot Journal of Educational The relationship between learning styles , gender and learning’, 17(1), pp. 56–67.

Fang, Y. *et al.* (2018) ‘Clustering the Learning Patterns of Adults with Low Literacy Skills Interacting with an Intelligent Tutoring System’, in *International Educational Data Mining Society*, p. 11. Available at: <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED593217&site=ehost-live&authtype=ip,uid>.

Fang, Z. *et al.* (2021) ‘The relationship between cognitive ability and BOLD activation across sleep–wake states’, *Brain Imaging and Behavior*.

Fatahi, S., Moradi, H. and Farmad, E. (2015) ‘Behavioral Feature Extraction To Determine Learning Styles in E-Learning Environments’, in *International Conference e-Learning*, pp. 66–72. Available at: <https://eric.ed.gov/?id=ED562499>.

Felder, R. M. (1988) ‘Learning and Teaching Styles in Engineering Education’, *engineering education*, 78(7), pp. 674–681. Available at: <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-1988.pdf> (Accessed: 11 December 2017).

Felder, R. M. and Spurlin, J. (2005) ‘Applications , Reliability and Validity of the Index of Learning Styles’, *International Journal of Engineering*, 21(1), pp. 113–112.

Felder, R. and Soloman, B. (2017) *Index of Learning Styles Questionnaire*. Available at: <https://www.webtools.ncsu.edu/learningstyles/> (Accessed: 15 December 2017).

Feng, Y., Iriarte, F. and Valencia, J. (2020) ‘Relationship between Learning Styles, Learning Strategies and Academic Performance of Chinese Students Who Learn Spanish as a Foreign Language’, *Asia-Pacific Education Researche*, 29(5). doi: <http://dx.doi.org/10.1007/s40299-019-00496-8>.

Ferguson, R. and Clow, D. (2017) ‘Where is the evidence? A call to action for learning analytics’, in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pp. 56–65.

Fischer, G. (2001) ‘User modeling in human-computer interaction’, *User Modeling and User-Adapted Interaction*, 11(1–2), pp. 65–86. doi: 10.1023/A:1011145532042.

Fleming, N. D. (2014) *The VARK Modalities*. Available at: <https://web.archive.org/web/20150314235648/http://vark-learn.com/introduction-to-vark/the->

vark-modalities/.

Gagne, E. (1985) *The cognitive psychology of school learning*. Boston: Little, Brown and Company.

García, P., Schiaffino, S. and Amandi, A. (2008) 'An enhanced Bayesian model to detect students' learning styles in Web-based courses', *Journal of Computer Assisted Learning*. Blackwell Publishing Ltd, 24(4), pp. 305–315. doi: 10.1111/j.1365-2729.2007.00262.x.

Gaševića, D., Kovanović, V. and Joksimović, S. (2017) 'Piecing the learning analytics puzzle: a consolidated model of a field of research and practice', *Learning: Research and Practice*, 3(1), pp. 63–78.

Gautam, D. *et al.* (2018) 'Automated Speech Act Categorization of Chat Utterances in Virtual Internships', *International Educational Data Mining Society*, pp. 341–347. Available at: [http://educationaldatamining.org/files/conferences/EDM2018/papers/EDM2018\\_paper\\_115.pdf](http://educationaldatamining.org/files/conferences/EDM2018/papers/EDM2018_paper_115.pdf).

Ghadirian, H., Fauzi Mohd Ayub, A. and Salehi, K. (2018) 'Students' perceptions of online discussions, participation and e-moderation behaviours in peer-moderated asynchronous online discussions', *Technology, Pedagogy and Education*. Routledge, 27(1), pp. 85–100. doi: 10.1080/1475939X.2017.1380695.

Glover, J., Bruning, R. and Filbeck, R. (1983) *Educational psychology principles and applications*. Boston: Little, Brown and Company.

Gonzalez, R. V. D. and Martins, M. F. (2017) 'Knowledge Management Process: a theoretical-conceptual research', *Gestao e Producao*, 24(2), pp. 248-265,.

Goodridge, W. H., Lawanto, O. and Santoso, H. B. (2017) 'A Learning Style Comparison between Synchronous Online and Face-to-Face Engineering Graphics Instruction', *International Education Studies*, 10(2), p. 1. doi: 10.5539/ies.v10n2p1.

Graf, Sabine and Kinshuk (2008) 'Analysing the behaviour of students in learning management systems with respect to learning styles', *Studies in Computational Intelligence*, 93, pp. 53–73. doi: 10.1007/978-3-540-76361\_3.

Graf, Sabine. and Kinshuk (2008) *Learner modelling through analyzing cognitive skills and learning styles*, *International Handbooks on Information Systems*. Springer Nature. doi: 10.1007/978-3-540-74155-8\_10.

Graf, S. and Kinshuk (2013) 'Dynamic Student Modelling of Learning Styles for Advanced Adaptivity in Learning Management Systems', *International Journal of Information Systems and Social Change*, 4(1), pp. 85–100. doi: 10.4018/jissc.2013010106.

Graf, S., Kinshuk and Liu, T.-C. (2009) 'Supporting Teachers in Identifying Students' Learning Styles in Learning Management Systems: An Automatic Student Modelling Approach', *Educational Technology & Society*, 12(4), pp. 3–14. doi: 10.1016/j.lindif.2010.11.016.

Graf, S. and Lin, T. (2008) 'IMPROVING STUDENT MODELING : THE RELATIONSHIP BETWEEN LEARNING STYLES AND COGNITIVE TRAITS', *Computers in Human*

*Behavior*, 24(2), pp. 122–137.

Graf, S. and Liu, T.-C. (2009) ‘Supporting Teachers in Identifying Students’ Learning Styles in Learning Management Systems: An Automatic Student Modelling Approach’, *Educational Technology & Society*, 12(4), pp. 3–14. Available at: [http://www.ifets.info/journals/12\\_4/2.pdf](http://www.ifets.info/journals/12_4/2.pdf) (Accessed: 11 December 2017).

Graf, S., Viola, S. R. and Kinshuk (2007) ‘AUTOMATIC STUDENT MODELLING FOR DETECTING LEARNING STYLE PREFERENCES IN LEARNING MANAGEMENT SYSTEMS \*’, in *IADIS International Conference Cognition and Exploratory Learning in Digital Age*.

Gross, R. (2012) *Psychology: The Science of Mind and Behaviour*. 6th edn. UK: Hodder Education.

Guzmán-Valenzuela, C. *et al.* (2021) ‘Learning analytics in higher education: a preponderance of analytics but very little learning?’, *International Journal of Educational Technology in Higher Education*. doi: <https://doi.org/10.1186/s41239-021-00258-x>.

Hamed, M. and Almabruk, A. (2021) ‘Perceptual Learning Style Preferences of English Major Libyan University Students and their Correlations with Academic Achievement’, *Advances in Language and Literary Studies*, 12(5), p. 1. doi: [10.7575/aiac.all.v.12n.5.p.1](https://doi.org/10.7575/aiac.all.v.12n.5.p.1).

Hampshire, A. *et al.* (2012) ‘Fractioning human intelligence.’, *Neuron*, 76(6), pp. 1225–1237.

Han, J., Kamber, M. and Pei, J. (2012) *Cluster Analysis: Basic Concepts and Methods, Data Mining: Concepts and Techniques*. doi: [10.1016/B978-0-12-381479-1.00010-1](https://doi.org/10.1016/B978-0-12-381479-1.00010-1).

Harrak, F., Bouchet, F. and Luengo, V. (2019) ‘From Student Questions to Student Profiles in a Blended Learning Environment’, *Journal of Learning Analytics*, 6(1), pp. 54–84. doi: [10.18608/jla.2019.61.4](https://doi.org/10.18608/jla.2019.61.4).

Hartigan, J. A. and Wong, M. (1979) ‘Algorithm AS 136: A K-Means Clustering Algorithm’, *Journal for Royal Statistical Society*, 28(1), pp. 100–108.

Hasibuan, M. S. *et al.* (2016) ‘A PROPOSED MODEL FOR DETECTING LEARNING STYLES BASED ON AGENT LEARNING’, *International Journal of Emerging Technologies in Learning*, 11(10), pp. 65–69. doi: [10.3991/ijet.v11i10.5781](https://doi.org/10.3991/ijet.v11i10.5781).

Helal, M. S. A., Ahmed, I. and Bhuiyan, M. E. M. (2021) ‘Impact of Education Management Information System (EMIS) on Teaching-Learning Development’, *International Journal of Academic Research in Progressive Education and Development*, 10(2), pp. 948 – 956. doi: <http://dx.doi.org/10.6007/IJARPED/v10-i2/10443>.

Herrmann, N. (1976) *The theory behind the HBDI and whole brain technology*.

Honey, P. and Mumford, A. (1982) *The Manual of Learning Styles*. Berkshire: Peter Honey.

*Ilias* (1998). Available at: [docu.ilias.de](http://docu.ilias.de) (Accessed: 21 December 2020).

Johnson, A. and Taatgen, N. (2005) ‘User Modeling’, *Handbook of human factors in Web*

design. Lawrence Erlbaum Associates.

Kalyuga (2012) 'Interactive distance education: A cognitive load perspective', *Journal of Computing in Higher Education*, 24(3), pp. 182 – 208.

Kamsa, I., Elouahbi, R. and El Khoukhi, F. (2018) 'Study Smart Not Hard', *Turkish Online Journal of Distance Education*, 19(1), pp. 62–74.

Kaplana, A. and Haenlein, M. (2019) 'Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence', *Business Horizons*, 62(1), pp. 15–25.

Kassambara, A. (2015) *Multivariate Analysis I: Practical Guide To Cluster Analysis in R*, Taylor & Francis Group. Available at:  
[https://pe56d.s3.amazonaws.com/o\\_1b64qp8fdsg81t4k1fongljq24m.pdf?AWSAccessKeyId=AKIAIP3NFJ5OR5FTFOQQ&Signature=ekwWQ2FpwaTjBwVMF20SsNGBSR4%3D&Expires=1522680996&response-content-disposition=inline;filename=%22preview.pdf%22&response-content-type=applicati](https://pe56d.s3.amazonaws.com/o_1b64qp8fdsg81t4k1fongljq24m.pdf?AWSAccessKeyId=AKIAIP3NFJ5OR5FTFOQQ&Signature=ekwWQ2FpwaTjBwVMF20SsNGBSR4%3D&Expires=1522680996&response-content-disposition=inline;filename=%22preview.pdf%22&response-content-type=applicati).

Kaufman, L. and Rousseeuw, J. P. (2005) *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons. Copyright.

Kay, J. and Kummerfeld, B. (2012) 'Lifelong learner modeling', in *Adaptive Technologies for Training and Education*. Cambridge, U.K: Cambridge University Press, pp. 140–167.

Khan, F. A. *et al.* (2019) 'Automatic student modelling for detection of learning styles and affective states in web based learning management systems', *IEEE Access*. IEEE, 7, pp. 128242–128262. doi: 10.1109/ACCESS.2019.2937178.

Kibuku, R. N., Ochieng, D. O. and Wausi, A. N. (2020) 'E-learning challenges faced by universities in Kenya: A literature review', *Electronic Journal of e-Learning*, 18(2), pp. 150–161. doi: 10.34190/EJEL.20.18.2.004.

Klein, S. (2012) *Learning: Principles and Applications*. 6th edn. SAGE.

Knight, S., Gibson, A. and Shibani, A. (2020) 'Implementing learning analytics for learning impact: Taking tools to task.', *Internet and Higher Education*, 45. doi: <https://doi.org/10.1016/j.iheduc.2020.100729>.

Kolb, D. (2015) *Experiential Learning: Experience as the Source of Learning and Development*. 2nd edn. Pearson Education.

Kroenke, D. and Boyle, R. (2019) *Using Management Information Systems*. 10th edn. Harlow, United Kingdom: Pearson Education.

Kurvinen, E. *et al.* (2020) 'Long term effects on technology enhanced learning: The use of weekly digital lessons in mathematics.', *Informatics in education*., 19(1), pp. 51–75.

Kuswara, H. *et al.* (2019) 'Stufflebeam's Model Application of Education Management Information Systems (EMIS) in Improving the Quality of Learning Services', *Nidhomul Haq*:



*Jurnal Manajemen Pendidikan Islam*, 6(1).

Landis, J. R. and Koch, G. G. (1977) 'The Measurement of Observer Agreement for Categorical Data', *Biometrics*, 33(1), p. 159. doi: 10.2307/2529310.

Latham, A. *et al.* (2012) 'A conversational intelligent tutoring system to automatically predict learning styles', *Computers & Education*. Pergamon, 59(1), pp. 95–109. doi: 10.1016/J.COMPEDU.2011.11.001.

Lee, S. Y. and Natriello, G. (2018) 'Identifying User Engagement Patterns in an Online Video Discussion Platform', in *Proceedings of the 11th International Conference on Educational Data Mining*, pp. 363–368.

Levine, B. *et al.* (2017) 'Cognitive and psychosocial function in retired professional hockey players.', *Journal of Neurology, Neurosurgery and Psychiatry*.

Lin, T.-Y. (2007) *Cognitive trait model for adaptive learning environments*. Massey University. Available at: <https://mro.massey.ac.nz/handle/10179/1451> (Accessed: 11 December 2017).

Lin, T. and Kinshuk (2004) 'Dichotomic node network and Cognitive Trait Model.', in *In: Proceedings of IEEE International Conference on Advanced Learning Technologies*. Los Alamitos, CA: IEEE Computer Science, pp. 702–704.

Lin, T., Kinshuk and Graf, S. (2007) 'Cognitive trait model and divergent associative learning', in *Proceedings - The 7th IEEE International Conference on Advanced Learning Technologies, ICAIT 2007*. IEEE Computer Society, pp. 325–327. doi: 10.1109/ICALT.2007.95.

Lindner, T., Puck, J. and Verbeke, A. (2020) 'Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies', *Journal of International Business Studies*. Palgrave Macmillan UK, 51(3), pp. 283–298. doi: 10.1057/s41267-019-00257-1.

Liu, H., He, H. and Qin, J. (2021) 'Does background sounds distort concentration and verbal reasoning performance in open-plan office?', *Applied Acoustics*, 172(15).

Liyanage, M. P. P., Gunawardena, K. S. L. and Hirakawa, M. (2014) 'Using Learning Styles to Enhance Learning Management Systems', *International Journal on Advances in ICT for Emerging Regions*, 07(07), pp. 1–10.

Liyanage, M. P. P., Gunawardena, K. S. L. and Hirakawa, M. (2016) 'Detecting Learning Styles in Learning Management Systems Using Data Mining', *Journal of Information Processing*, 24(4), pp. 740–749. doi: 10.2197/ipsjjip.24.740.

Lurquin, J. and Michaelson, L. (2017) *Outcome Task: Operation Span (OSpan)*, *OSFHOME*. Available at: <https://osf.io/zwk7r/>.

Magoulas, G. D. and Dimakopoulos, D. N. (2005) 'Designing Personalised Information Access to Structured Information Spaces.', in *Proceedings of the 1st International Workshop Workshop on New Technologies for Personalized Information Access*.

- Makokha, G. L. and Mutisya, D. N. (2016) 'Status of E-learning in public Universities in Kenya', *International Review of Research in Open and Distance Learning*, 17(3), pp. 341–359. doi: 10.19173/irrodl.v17i3.2235.
- Miller, G. . (1956) 'The magic number seven, plus or minus two: Some limit of our capacity for processing information', *Psychological Review*, 63, pp. 81–96.
- Mitchell, T. M. (1997) *Machine Learning*. New York, USA: McGRAW-HILL.
- Montaner, M., López, B. and De La Rosa, J. L. (2003) 'A Taxonomy of Recommender Agents on the Internet', *Artificial Intelligence*, 19(4), pp. 285–330.
- Moodle (2018). Available at: moodle.org (Accessed: 21 December 2020).
- Moraes, T. C. *et al.* (2020) 'LSBCTR: A Learning Style-Based Recommendation Algorithm.', in *IEEE Frontiers in Education Conference (FIE)*, pp. 1–9.
- Mukhtar, M., Fransiska, A. B. and Wahyudi, M. (2018) 'Management information systems doctorate program of educational management (DOCPem)', in *6th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–5. doi: <https://doi.org/10.1109/CITSM.2018.8674314>.
- Myers, I. B. (1995) *Gifts Differing: Understanding Personality Type*. Reprint. Davies-Black Publishing.
- Naeim, M. *et al.* (2022) 'Prediction of academic performance based on learning style and critical thinking among medical students', *Journal of Pedagogical Research*, 6(1), pp. 57–66. doi: 10.33902/jpr.2022175483.
- Nafea, S. M., Siewe, F. and He, Y. (2019) 'A novel algorithm for course learning object recommendation based on student learning styles', in *International Conference on Innovative Trends in Computer Engineering (ITCE)*.
- Namoun, A. and Alshantqiti, A. (2020) 'Predicting student performance using data mining and learning analytics techniques: A systematic literature review.', *Applied Sciences*. doi: <https://doi.org/10.3390/app11010237>.
- Ozdemir, A. *et al.* (2018) 'An Intelligent System for Determining Learning Style', *International Journal of Research in Education and Science*, 4(1), pp. 208–214. doi: 10.21890/ijres.383140.
- Paas, F. and Ayres, P. (2014) 'Cognitive load theory: A broader view on the role of memory in learning and education.', *Educational Psychology Review*, 26(2), pp. 191 – 195. doi: 10.1007/s10648-014-9263-5.
- Paivio, A. (1969) 'Mental Imagery in Associative Learning and Memory', *Psychological Review*, 76(3), pp. 241 – 263.
- Paivio, A. and Desrochers, A. (1981) 'Mnemonic techniques in second language learning', *Journal of Educational Psychology*, 73, pp. 780–795.
- Park, S. (2017) 'Analysis of time-on-task, behavior experiences, and performance in two online

courses with different authentic learning tasks’, *International Review of Research in Open and Distance Learning*, 18(2), pp. 213–233. doi: 10.19173/irrodl.v18i2.2433.

Patnaik, A. K., Bhuyan, P. K. and Krishna Rao, K. V. (2016) ‘Divisive Analysis (DIANA) of hierarchical clustering and GPS data for level of service criteria of urban streets’, *Alexandria Engineering Journal*. Faculty of Engineering, Alexandria University, 55(1), pp. 407–418. doi: 10.1016/j.aej.2015.11.003.

Perina, K. *et al.* (2021) ‘Mental Speed Test - Version 1’, *Psychology Today*. Available at: <https://www.psychologytoday.com/us/tests/iq/mental-speed-test-version-1>.

Pham, Q. D. and Florea, A. M. (2013) ‘A Method for Detection of Learning Styles in Learning Management Systems’, *UPB Scientific Bulletin, Series C: Electrical Engineering*, 75(4), pp. 3–12.

Piao, G. and Breslin, J. G. (2018) ‘Inferring user interests in microblogging social networks: a survey.’, *User Modeling and User-Adapted Interaction*, 28(3), pp. 277–329.

Poellhuber, B. and Roy, N. (2019) ‘Understanding participant’s behaviour in massively open online courses’, *International Review of Research in Open and Distance Learning*, 20(1), pp. 222–242.

Rajper, S. *et al.* (2016) ‘Automatic Detection of Learning Styles on Learning Management Systems using Data Mining Technique’, *Indian Journal of Science and Technology*, 9(15). doi: 10.17485/ijst/2016/v9i15/85959.

Renato, R. M. I. (2020) ‘Adaptive Virtual Learning Environment based on Learning Styles for Personalizing E-learning System: Design and Implementation’, *International Journal of Recent Technology and Engineering (IJRTE)*, 8(6), pp. 3398–3406. doi: 10.35940/ijrte.f8901.038620.

Rets, I. *et al.* (2021) ‘Exploring critical factors of the perceived usefulness of a learning analytics dashboard for distance university students.’, *International Journal of Educational Technology in Higher Education*. doi: <https://doi.org/10.1186/s41239-021-00284-9>.

Riding, R. and Rayner, S. (1998) *Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behavior*. London: David Fulton Publisher.

Rojas, R. (1996) *Neural Networks: A Systematic Introduction*. Berlin: Springer.

Romero, C. *et al.* (2007) ‘Personalized Links Recommendation Based on Data Mining in Adaptive Educational Hypermedia Systems’, in Erik Duval, Ralf Klamma, M. W. (ed.) *Creating New Learning Experiences on a Global Scale. Second European Conference on Technology Enhanced Learning, EC-TEL 2007, Crete, Greece, September 17-20, 2007*. Berlin Heidelberg: Springer, pp. 292 – 306.

Romero, C. and Ventura, S. (2020) ‘Educational data mining and learning analytics: An updated survey’, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3).

Rothrock, L. *et al.* (2002) ‘Review and Reappraisal of Adaptive Interfaces: Toward Biologically-

Inspired Paradigms.’, in *Theoretical Issues in Ergonomic Science*.

Ruiji, L. (2012) ‘The development on multimedia teaching resources based on information processing theory.’, *International Journal of Advancements in Computing Technology*, 4(2), pp. 58 – 64.

Rumelhart, D. E. and McClelland, J. L. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. Cambridge, Massachusetts: MIT Press.

Salomon, G. and Perkins, D. (1998) ‘Individual and social aspects of learning’, *Review of Research in Education*, 23, pp. 1–24.

Sandbhor, S. and Chaphalkar, N. B. (2018) ‘Impact of Outlier Detection on Neural Networks Based Property Value Prediction’, in Satapathy, S. et al. (eds) *Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing*.

Seidenberg, M. and McClelland, J. L. (1989) ‘A distributed, developmental model of word recognition and naming’, *Psychological Review*, 96, pp. 523 – 568.

Sensuse, D. I., Hasani, L. M. and Bagustari, B. (2020) ‘Personalization Strategies Based on Felder-Silverman Learning Styles and Its Impact on Learning: A Literature Review’, in *3rd International Conference on Computer and Informatics Engineering (IC2IE)*, pp. 293–298.

Shah, V. (2013) ‘Impact of management information systems (MIS) on school administration: What the literature says’, in *5th World Conference on Educational Sciences*, pp. 2799 – 2804.

Sim, J. and Wright, C. C. (2005) ‘The kappa statistic in reliability studies: Use, interpretation, and sample size requirements’, *Physical Therapy*, 85(3), pp. 257–268. doi: 10.1093/ptj/85.3.257.

Slate, J. R. and Charlesworth, J. R. (1989) ‘Information Processing Theory: Classroom Applications.’, *Educational Resources Information Center*, pp. 1–17.

Smolensky, P. (1999) ‘Grammar-based Connectionist Approaches to Language’, *Cognitive Science*, 23(4), pp. 589 – 613.

Sridharan, S. et al. (2021) ‘Adaptive Learning Management Expert System with Evolving Knowledge Base and Enhanced Learnability’, *Education and Information Technologies*, 26(5), pp. 5895–5916.

Sweet, L. H. (2011) ‘Information Processing Speed’, *Encyclopedia of Clinical Neuropsychology*. Springer - New York.

Sweta, S. and Lal, K. (2016) ‘Learner Model for Automatic Detection of Learning Style Using FCM in Adaptive E-Learning System’, *IOSR Journal of Computer Engineering*, 18(2), pp. 18–24. doi: 10.9790/0661-1802041824.

Tarus, J. K. and Gichoya, D. (2015) ‘E-learning in Kenyan universities: Preconditions for successful implementation’, *Electronic Journal of Information Systems in Developing Countries*, 66(1). doi: 10.1002/j.1681-4835.2015.tb00474.x.

- Taylor, C. (2013) ‘Cognitive load theory: Sometimes less is more.’, *Journal on School Educational Technology*, 96(1), pp. 61 – 68.
- Triantafyllou, E., Georgiadou, E. and Economides, A. (2009) ‘The role of user model in CAT: Exploring adaptive variables’, *Technology, Instruction, Cognition and Learning (TICL)*, 5(1), pp. 69–89. Available at: [http://www.edutech.gr/public/user\\_model\\_CAT.pdf](http://www.edutech.gr/public/user_model_CAT.pdf).
- Turner, M. L. and Engle, R. W. (1989) ‘Is working memory capacity task dependant ?’, *Journal of Memory and Language*, pp. 127–154. doi: 10.1016/0749-596X(89)90040-5.
- Ugwude, A. C. and Ugwude, D. I. (2020) ‘Challenges of Education Management Information Systems on Primary School Administration in Nsukka Local Government Education Authority of Enugu State’, *Journal of Educational Leadership*, 5(2), pp. 49–56.
- Ulker, D. and Yilmaz, Y. (2016) ‘Learning management systems and comparison of open source learning management systems and proprietary learning management systems’, *Journal of Systems Integration*, 7(2), pp. 18–24.
- Umer, R. *et al.* (2021) ‘Current stance on predictive analytics in higher education: Opportunities, challenges and future directions.’, *Interactive Learning Environments*, <https://doi.org/10.1080/10447199.2021.1911111>.
- Vaerenbergh, S. Van and Perez-Suay, A. (2018) ‘Intelligent Learning Management Systems : Overview and Application in Mathematics Education’, in Richard, P. R., V´elez, M. P., and Van Vaerenbergh, S. (eds) *Mathematics Education in the Age of Artificial Intelligence: How Artificial Intelligence can Serve Mathematical Human Learning*. Cham: Springer International Publishing, pp. 1–15.
- Vaerenbergh, S. Van and Perez-Suay, A. (2022) ‘A classification of artificial intelligence systems for mathematics education.’, in Richard, P. R., V´elez, M. P., and Van Vaerenbergh, S. (eds) *Mathematics Education in the Age of Artificial Intelligence: How Artificial Intelligence can Serve Mathematical Human Learning*. Cham: Springer International Publishing, pp. 89–106.
- Valacich, J. and Schneide, C. (2015) *Information Systems Today: Managing in the Digital World (7th Edition) Information Systems Today: Managing in the Digital World*. 7th edn. Harlow, United Kingdom: Palgrave Macmillan UK.
- Varlamis, I. and Apostolakis, I. (2006) ‘The Present and Future of Standards for E-Learning Technologies’, *Interdisciplinary Journal of Knowledge and Learning Objects*, 2, pp. 59–76. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.98.9621>.
- Villegas-Ch, W. *et al.* (2018) ‘Analysis of web-based learning systems by data mining’, in *2017 IEEE 2nd Ecuador Technical Chapters Meeting, ETCM 2017*, pp. 1–5. doi: 10.1109/ETCM.2017.8247553.
- Vinney, C. (2020) *Information Processing Theory: Definition and Examples, ThoughtCo: Science, Maths and Social sciences*. Available at: <https://www.thoughtco.com/information-processing-theory-definition-and-examples-4797966>.
- Waheed, H. *et al.* (2020) ‘Predicting academic performance of students from VLE big data using deep learning models’, *Computers in Human Behavior*, 104.

- Wason, P. . (1960) 'On the failure to eliminate hypotheses in a conceptual task', *The Quarterly Journal of Experimental Psychology*, 12, pp. 129–140.
- Watkins, C. and Mortimore, P. (1999) *Pedagogy: What do we Know?". Understanding Pedagogy and its Impact on Learning.*
- Williams, K. B. and Sawyer, C. S. (2014) *Using information technology*. 11th edn. New York, New York, USA: McGRAW-HILL.
- Wilson, A. *et al.* (2017) 'Learning analytics: Challenges and limitations.', *Teaching in Higher Education*, 22(8). doi: <https://doi.org/10.1080/10571725.2017.1380000>.
- Van Wingerden, C. (2021) 'Designing for Inclusion within the Learning Management System: Social Justice, Identities, and Online Design for Digital Spaces in Higher Education', *International Journal of Educational and Pedagogical Sciences*, 15(8), pp. 684–692.
- Woolfolk, A. E. (2004) *Educational psychology*. Boston: 9: Allyn & Bacon.
- Yen, C.-H. *et al.* (2015) 'An Analytics-Based Approach to Managing Cognitive Load by Using Log Data of Learning Management Systems and Footprints of Social Media', *Educational Technology & Society*, 18(4), pp. 141–158.
- Yu, C.-H., Wu, J. and Liu, A.-C. (2019) 'Predicting Learning Outcomes with MOOC Clickstreams', *Education Sciences*, 9(104). Available at: [www.mdpi.com/journal/education](http://www.mdpi.com/journal/education).
- Zeglen, E. and Rosendale, J. A. (2018) 'Increasing online information retention: Analysing the effects of visual hints and feedback in educational games', *Journal of Open, Flexible and Distance Learning*, 22(1), pp. 22–33.
- Zheng, A. (2015) *Evaluating Machine Learning Models*. First Edit. O'Reilly Media, Inc.



# APPENDICES

## *Appendix 1. Publications*

### **Journal Article**

Lwande, C., Muchemi, L. and Oboko, R. (2021) 'Identifying Learning Styles and Cognitive Traits in a Learning Management System', *Heliyon*. Elsevier, 7(May 2019), p. e07701. DOI: 10.1016/j.heliyon.2021.e07701.

Lwande, C., Oboko, R. & Muchemi, L. Learner Behavior prediction in a Learning Management system. *Education and Information Technologies* (2020). <https://doi.org/10.1007/s10639-020-10370-6>

### **Conference (Publisher: IEEE)**

Lwande, C., Oboko, R. & Muchemi, "Behavior Prediction in a Learning Management System," *2019 IST-Africa Week Conference (IST-Africa)*, Nairobi, Kenya, 2019, pp. 1-10, DOI: 10.23919/ISTAFRICA.2019.8764877.

### **Book Chapters**

Lwande, C., Muchemi, L. and Oboko, R. (2019a) 'Behavior Modeling in a Learning Management System', in Kwanya, T. et al. (eds) *Digital Technologies for Information and Knowledge Management*. 2ND ed. Nairobi: The Technical University of Kenya •, pp. 191–198.



Appendix 2. Letter of Request to Conduct Research at the University



**UNIVERSITY OF NAIROBI**  
**OFFICE OF THE DEPUTY VICE - CHANCELLOR**  
(Research, Production & Extension)  
Prof. Lucy W. Irungu B.Sc., M.Sc., Ph.D.

P.O. Box 30197-GPO,  
00100, Nairobi-Kenya  
Telephone: +254-20-2315416 (DI), 318262

Fax: 0202317251  
Email: dvrpe@uonbi.ac.ke

UON/RPE/3/5/Vol.XVI

April 20, 2016

Charles L. Omondi  
University of Nairobi  
ICTC Center  
P.O. Box 30197 – 00100  
NAIROBI

Dear Omondi,

**AUTHORITY TO CARRY OUT RESEARCH AT THE UNIVERSITY OF NAIROBI**

I refer to your request dated April 4, 2016 to conduct research at the University of Nairobi, for your PhD Degree entitled: “*Student Behaviour modeling in a Learning Management System*”.

I write to inform you that your request has been approved.

You are however required to share the findings of your study with the University of Nairobi by depositing a copy of your research findings with the Director, Library and Information Services on completion of your study.

**PROF. HENRY W. MUTORO**  
**AG. DEPUTY VICE-CHANCELLOR**  
**(RESEARCH PRODUCTION AND EXTENSION)**

c.c. Vice-Chancellor  
DVC, A&F  
DVC, AA  
DVC, SA  
Director, Library & Information Services  
Director, ICTC



ISO 9001:2008 CERTIFIED

*The Fountain of Knowledge Providing leadership in academics excellence*

## Appendix 3. The Electronic Learning Module Sections hosted on Claroline LMS

(source: University of Nairobi LMS)

### Module 1: Psychology Theory eLearning book

- HCH 100: BEHAVIOURAL SCIENCES
  - ▣ HCH 100: BEHAVIOURAL SCIENCES
    - ▣ Module 1: Fundamentals of Psychology
      - ▣ Unit I: Basic Concepts of Psychology
        - ▣ 1.1: Objectives
        - ▣ SECTION 1: Introduction to Psychology
          - ▣ 1.1 Section Introduction
          - ▣ 1.2 Section Objectives
          - ▣ 1.3 The Science of Psychology
          - ▣ 1.4 Biological Basis of Behaviour
          - ▣ 1.5 Sensation and Perception
          - ▣ 1.6 Section Summary
        - ▣ SECTION 2: Human Intelligence
          - ▣ 2.1: Section 2 Introduction
          - ▣ 2.2 Section 2 Objectives
          - ▣ 2.3 Human Intelligence
          - ▣ 2.4 Approaches to intelligence
          - ▣ 2.5 Assessing intelligence/IQ testing
          - ▣ 2.6: Human Intelligence Activity
          - ▣ 2.7 Summary
          - ▣ 2.8 References
          - ▣ 2.9 Self-test questions
        - ▣ SECTION 3: Learning Theories
          - ▣ 3.1 Section 3 Introduction
          - ▣ 3.2 Section 3 Objectives
          - ▣ 3.3 Learning
          - ▣ 3.4 Theories of learning

## Appendix 4. A Sample Online SCORM Content on Claroline LMS

(source: university of Nairobi LMS)

### Module 1: Psychology Theory eLearning book

- HCH 100: BEHAVIOURAL SCIENCES
  - ▣ HCH 100: BEHAVIOURAL SCIENCES
    - ▣ Module 1: Fundamentals of Psychology
      - ▣ Unit I: Basic Concepts of Psychology
        - ▣ 1.1: Objectives
        - ▣ SECTION 1: Introduction to Psychology
          - ▣ 1.1 Section Introduction
          - ▣ 1.2 Section Objectives
          - ▣ 1.3 The Science of Psychology
          - ▣ 1.4 Biological Basis of Behaviour
          - ▣ 1.5 Sensation and Perception
          - ▣ 1.6 Section Summary
        - ▣ SECTION 2: Human Intelligence
          - ▣ 2.1: Section 2 Introduction
          - ▣ 2.2 Section 2 Objectives
          - ▣ 2.3 Human Intelligence
          - ▣ 2.4 Approaches to intelligence
          - ▣ 2.5 Assessing intelligence/IQ testi...
          - ▣ 2.6: Human Intelligence Activity
          - ▣ 2.7 Summary
          - ▣ 2.8 References
          - ▣ 2.9 Self-test questions
        - ▣ SECTION 3: Learning Theories
          - ▣ SECTION 4: Memory
          - ▣ SECTION 5: Language Development
          - ▣ SECTION 6: Motivation and Emotion
          - ▣ Unit 1: Self Test Questions
          - ▣ UNIT 2: DEVELOPMENTAL PSYCHOLO...
          - ▣ Section 1: Human Development, Part ...
          - ▣ Section 2: Human Development II: Fr...

**Unit I: Basic Concepts of Psychology**

---

**Introduction**

Welcome to this module on Fundamentals of psychology. Unit one introduces you to psychology and the basic concepts in psychology. We hope that you will enjoy learning about human behaviour and psychology. We are all in an everyday sense 'researchers', trying to make sense of our world and we long to emerge like Archimedes shouting "Eureka! Eureka!" When I was a young student, whenever I saw "Christ is the answer" sign post on the church on Valley Road, I often wondered "what is the question?" Anyways as I navigated through my journey of life and tackled life's great lessons, I now, kind of know the questions! Similarly, one could say, in a scientific sense - "psychology is the answer" in that, psychology endeavors to make sense of our world through the scientific study of the mind and behavior.

In this Unit we will explore the basic concepts that lay the foundations of psychology. We also introduce you to sub-specialities within psychology, orientate you to various perspectives and key debates in psychology, provide a bio-psycho-social framework to understand basis of human behaviour. This unit is divided into six sections:

1. Section 1: Introduction to Psychology
- 1.2 The Science of Psychology (nature and its definition)
- 1.3 The Biology of Behavior (how neurons transmit signals; basic neuroanatomy (functions of different parts of the brain)
- 1.4 Sensation and Perception (the neural basis of sensory processes, such as vision, hearing, smell and taste; movement and autonomic functions).
2. Section 2: Intelligence
3. Section 3: Learning
4. Section 4: Memory

*Appendix 5. The Content Navigation Statistics generated from Claroline LMS logs*  
 (source: University of Nairobi LMS)

- Name :
- First name : JEPCHUMBA
- Email : sharonne @gmail.com

Module	Last session time	Total time	Module status	Progress
HCH 100: BEHAVIOURAL SCIENCES				0%
Module 1: Fundamentals of Psychology	00:00:35	00:00:35	completed	100%
Unit 1: Basic Concepts of Psychology	00:00:25	00:00:25	completed	100%
1.1: Objectives				0%
SECTION 1: Introduction to Psychology	00:00:24	00:00:24	completed	100%
1.1 Section Introduction	00:01:10	00:01:10	completed	100%
1.2 Section Objectives				0%
1.3 The Science of Psychology	00:00:24	00:00:24	completed	100%
1.4 Biological Basis of Behaviour	00:03:38	00:03:38	completed	100%
1.4a Biological Basis of Behaviour cont...	00:02:58	00:02:58	completed	100%

*Appendix 6. Sample MCQ hosted on Claroline LMS (source: The University of Nairobi LMS)*

**Exercise : BEHAVIOURAL SCIENCES MID SEMESTER 1 TEST (MCQS)1**

- Maximum allowed time : 2 h.
- Attempt 1 on 2
- Available from November 24, 2017 at 12:58 PM until November 24, 2017 at 06:00 PM

**Question 1**

**47. What is Communication?**

- Utterance of words
- Speaking to people
- Exchange of thoughts, messages, or information, as by speech, signal, or writing
- Conveying the message
- Conveying the thoughts

Multiple choice (Unique answer)

**Question 2**

**26. The defense mechanisms**

- Emerge randomly
- Help individuals cope with their internal and external anxiety and distress
- Cannot be brought under conscious control to ward off anxiety
- Operate to maintain sense of serenity
- Do not contribute to the formation of personality traits

*Appendix 7. Sample Mark Sheet generated from Claroline LMS*

(source: University of Nairobi LMS)

user_id	Worst score	Best score	Average score	Attempts	Average Time
280780	5	58	31.5	2	1 h. 1 sec.
280816	6	60	33	2	52 min. 49 sec.
285803	2	64	33	2	44 min. 24 sec.
280765	0	67	33.5	2	1 h. 45 sec.
289276	1	67	34	2	49 min. 31 sec.
285582	0	71	35.5	2	57 min. 13 sec.
267075	37	37	37	1	20 min. 40 sec.

*Appendix 8. Sample Raw Data from LMS Logs*

(Source: University of Nairobi LMS).

Sample data for modules' content navigation. The figures show the number of pages viewed.

	UNIT 1	SEC 1	1.1	1.2 - 1.8	2.1	2.2-2.5	2.6-sum
280774	100	100	100	100	0	0.5	100
280775	100	100	100	100	100	0.7	0
280776	100	100	100	0.4	0	0.6	100
280777	0	0	100	0.8	0	0.8	0
280778	100	100	100	100	100	100	100
280779	100	100	100	100	100	100	100
280780	100	100	100	100	0	0.7	100
280781	100	100	100	0.6	100	0.7	100
280782	100	100	100	0.4	0	0.8	100
280783	0	100	100	0.8	100	0.7	100
280784	100	100	100	100	100	100	100
280785	100	100	100	100	100	1	100
280786	100	100	100	100	100	0.9	100
280787	0	0	100	0.6	0	0.5	100
280788	100	100	0	0.6	100	0.5	100
280789	100	100	100	0.8	0	0.9	0
280790	100	100	100	100	100	0.9	100
280793	100	100	100	100	100	1	100

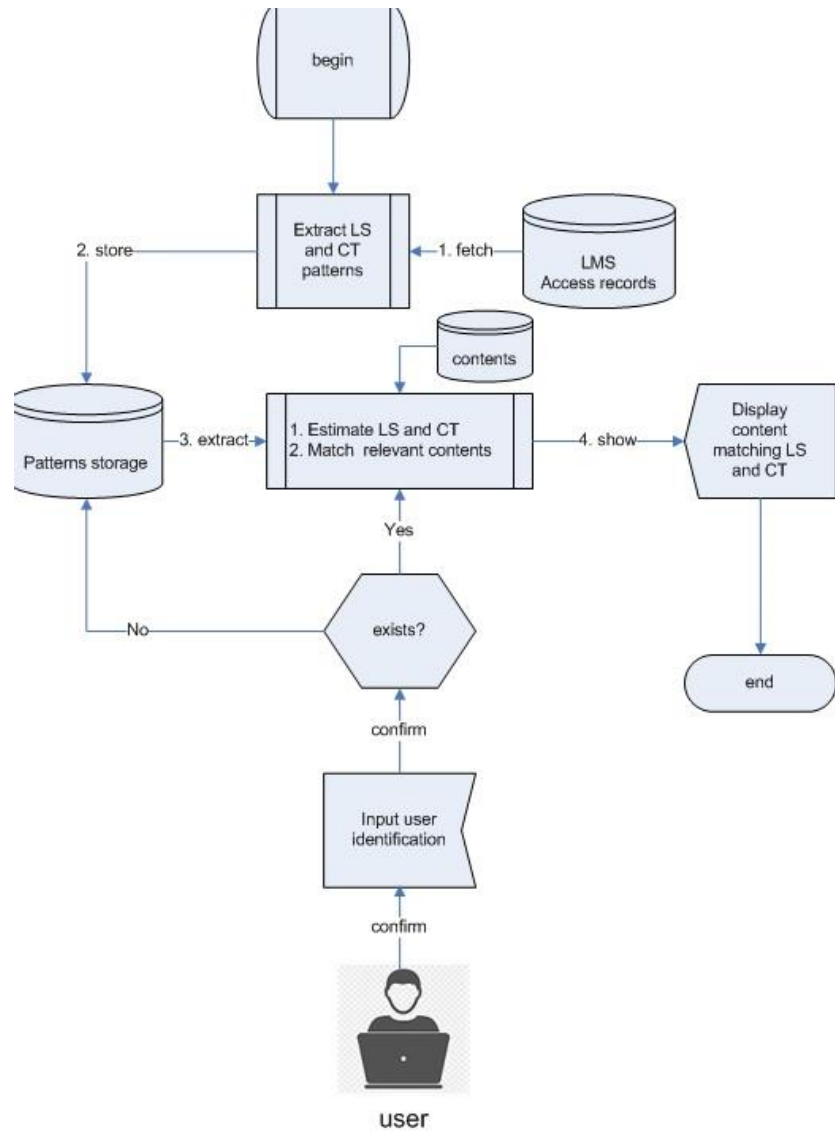
Sample raw data for time spent on module pages. The figures indicate time spent in minutes.

student_id	unit 1	sec 1	1.3	1.4-1.6	1.7-sum	sec-2	2.3	2.4	2.5-sum
280774	6	24	9	48	0	7	66	9	11
280775	7	5	5	35	5	25	70	6	7
280776	7	12	9	46	10	13	142	24	16
280777	5	4	5	143	27	32	691	52	9
280778	7	8	16	65	6	23	107	10	52
280779	7	5	7	26	8	7	50	10	10
280785	0	101	0	72	7	17	71	8	8
280790	30	20	25	241	176	52	643	10	7
280791	7	11	41	31	6	12	108	55	9
280792	4	6	3	39	5	0	52	0	6
280793	49	322	9	36	5	0	103	8	7
280794	0	0	0	47	27	21	220	0	0

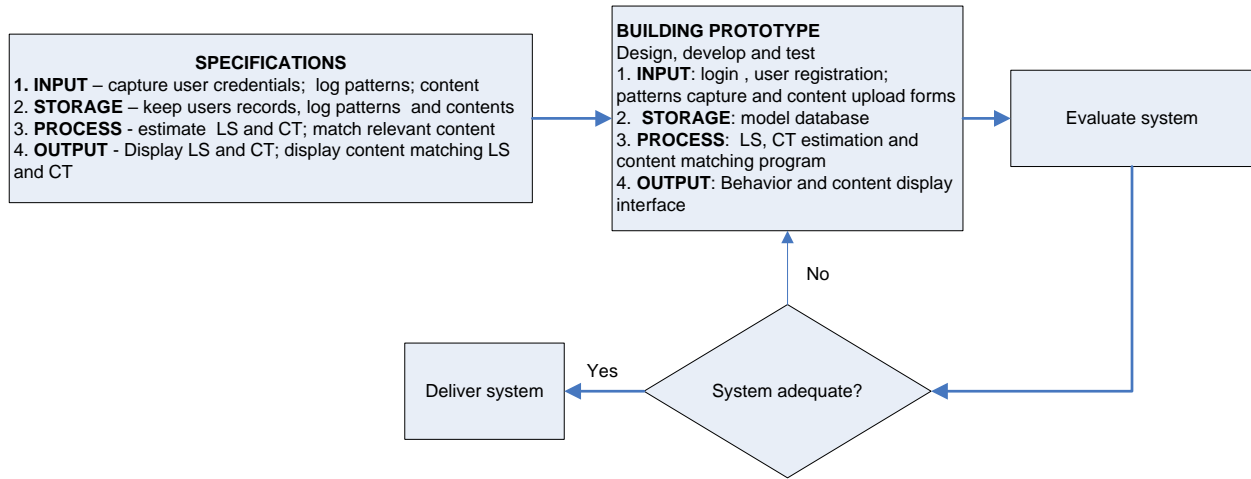
Sample raw data for the mark sheet. The figures indicate marks scored in online tests graded out of 100.

user_id	Worst score	Best score	Average score	Attempts	Average Time
	0	0	0	1	2 min. 26 sec.
280774	0	5	2	3	50 sec.
280775	3	59	31	2	25 min. 59 sec.
280780	5	58	31.5	2	1 h. 1 sec.
280816	6	60	33	2	52 min. 49 sec.
285803	2	64	33	2	44 min. 24 sec.
280765	0	67	33.5	2	1 h. 45 sec.
289276	1	67	34	2	49 min. 31 sec.
285582	0	71	35.5	2	57 min. 13 sec.
267075	37	37	37	1	20 min. 40 sec.
291189	4	70	37	2	52 min. 33 sec.
289278	38	38	38	1	22 min. 19 sec.
291033	1	75	38	2	18 min. 18 sec.
280812	39	39	39	1	17 min. 39 sec.
276467	40	40	40	1	2 h. 4 min. 26 sec.
285736	13	67	40	2	50 min. 5 sec.
280792	18	62	40	2	7 min. 54 sec.
291119	14	68	41	2	52 min. 20 sec.
291187	6	80	43	2	12 min. 30 sec.
283397	43	43	43	1	13 min. 4 sec.
285408	1	69	43.67	3	43 min. 53 sec.

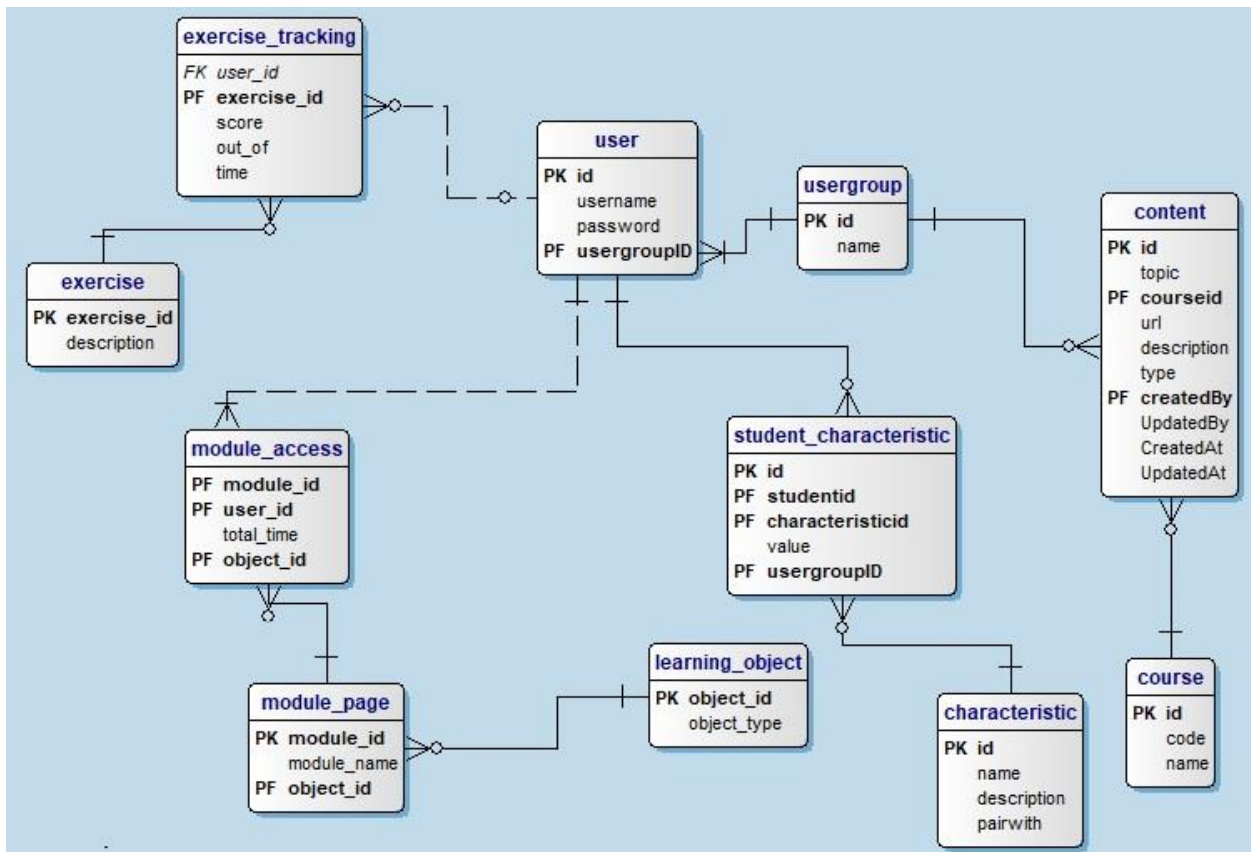
Appendix 9. The System Flow Diagram for Cognilearn Model



## Appendix 10. Evolutionary Prototyping Process



## Appendix 11. Model Entity-Relationship Diagram (ERD for the Cognilearn Model Database.



*Appendix 12. Data Dictionary for the CogniLearn Model Database.*

Table	Description of records	Fields
Users	Details of users e.g. student, tutor enrolled in a course	id, username, password, usergroupID
Usergroup	The role assigned to a user e.g. student, tutor, administrator	Id, name
Characteristic	Details of LS and CT	Id, name, description, pairwith
student_characteristic	LS and CT associated with each student	Id, studentid, characteristic, value
course	Adaptive course	Id, code, name
Content	Adaptive course contents	Id, topic, coursed, URL, description, type
Module page	Content pages of electronic books hosted in LMS	Module_id, name, or title
Module accessed	Content pages visited by students	Mudule_id, user_id of a student who accessed, duration
Learning object	Type of content page visited by a student	Module_id, content type for example outline, introduction, detailed notes, illustrations, summary
Exercise	Tests created in the LMS	Exercise id, title
Exercise tracking	Mark sheet or results and time taken	exercise_id, user_id of a student who attempted, marks, duration

*Appendix 13. CogniLearn Model Program Code Segments*

Sample codes for implementing learning styles and cognitive traits estimation functions

Estimating associative learning ability.

```
select
user_id,linear_pages_ratio,
repeated_pages_pages_ratio,repeated_time_ratio,linear_time_ratio,

(linear_pages_ratio*linear_time_ratio)+((1-repeated_pages_pages_ratio)*(1-repeated_time_ratio)) as linear_p,

(repeated_pages_pages_ratio*repeated_time_ratio)+((1-linear_pages_ratio)*(1-linear_time_ratio)) as repeat_p,

(linear_pages_ratio*linear_time_ratio)+((1-repeated_pages_pages_ratio)*(1-repeated_time_ratio)) +

(repeated_pages_pages_ratio*repeated_time_ratio)+((1-linear_pages_ratio)*(1-linear_time_ratio)) as total_p

from vw_CT_ALA_patterns_repeat_linear
```

**Estimating information processing speed**

```
select
user_id,regno,time_on_detailed_pages_ratio,detailed_pages_ratio,time_on_exercise_ratio,exercise_score_ratio,

((1-time_on_detailed_pages_ratio)*(detailed_pages_ratio)) +
((1-time_on_exercise_ratio)*(exercise_score_ratio)) as High,
((time_on_detailed_pages_ratio)*(1-detailed_pages_ratio)) +
((time_on_exercise_ratio)*(1-exercise_score_ratio)) as Low,
```



```

(((1-time_on_detailed_pages_ratio)*(detailed_pages_ratio)) +
((1-time_on_exercise_ratio)*(exercise_score_ratio)) )
+
(
((time_on_detailed_pages_ratio)*(1-detailed_pages_ratio)) +
((time_on_exercise_ratio)*(1-exercise_score_ratio))
)as total
from lb_CT_IPS_IRA_patterns

```

### Estimating inductive reasoning ability

#### **select**

```
user_id,regno,time_on_revision,revision_pages_ratio,time_on_exercise_ratio,exercise_score_ratio,
```

```

((time_on_revision)*(revision_pages_ratio)) +
((time_on_exercise_ratio)*(exercise_score_ratio)*(exercises_ratio)) as High,

```

```

((1-time_on_revision)*(1-revision_pages_ratio)) +
((1-time_on_exercise_ratio)*(1-exercise_score_ratio)*(1-exercises_ratio)) as Low,

```

```

((time_on_revision)*(revision_pages_ratio)) +
((time_on_exercise_ratio)*(exercise_score_ratio)*(exercises_ratio))
+
((1-time_on_revision)*(1-revision_pages_ratio)) +
((1-time_on_exercise_ratio)*(1-exercise_score_ratio)*(1-exercises_ratio))
as total
from lb_CT_IPS_IRA_patterns

```

### Estimating working memory capacity

#### **select**

```

user_id,regno,global_pages_ratio as nonlinear_pages,sequential_pages_ratio as linear_pages,
global_pages_time_ratio as nonlinear_time,sequential_pages_time_ratio as linear_time,
global_pages_ratio/(global_pages_ratio+sequential_pages_ratio) as nonlinear_pages_ratio,
sequential_pages_ratio/(global_pages_ratio+sequential_pages_ratio) as linear_pages_ratio,
global_pages_time_ratio/(global_pages_time_ratio+sequential_pages_time_ratio) as nonlinear_time_ratio,
sequential_pages_time_ratio/(global_pages_time_ratio+sequential_pages_time_ratio) as linear_time_ratio

```

```
from vw_LS_sequential_global_patterns_p
```

### Estimating active reflective learning style

#### **select**

```

user_id,username,time_on_detailed_pages_ratio,time_on_revision,time_on_outline_ratio,time_on_exercise_ratio,
detailed_pages_ratio,outline_pages_ratio,revision_pages_ratio,exercises_rati,

```

```

((1-time_on_detailed_pages_ratio)*(1-detailed_pages_ratio))+((exercises_ratio)*(1-time_on_exercise_ratio))+
((time_on_revision)*(revision_pages_ratio)) as active,

```

$((\text{time\_on\_detailed\_pages\_ratio}) * (\text{detailed\_pages\_ratio})) + ((1 - \text{exercises\_ratio}) * (\text{time\_on\_exercise\_ratio})) + ((\text{time\_on\_revision}) * (1 - \text{revision\_pages\_ratio}))$  **as** reflective,

$((1 - \text{time\_on\_detailed\_pages\_ratio}) * (1 - \text{detailed\_pages\_ratio})) + ((\text{exercises\_ratio}) * (1 - \text{time\_on\_exercise\_ratio})) + ((\text{time\_on\_revision}) * (\text{revision\_pages\_ratio}))$   
+

$((\text{time\_on\_detailed\_pages\_ratio}) * (\text{detailed\_pages\_ratio})) + ((1 - \text{exercises\_ratio}) * (\text{time\_on\_exercise\_ratio})) + ((\text{time\_on\_revision}) * (1 - \text{revision\_pages\_ratio}))$  **as** total

Estimating sensing intuitive learning style

**select**

user\_id, username, time\_on\_detailed\_pages\_ratio, time\_on\_revision, time\_on\_outline\_ratio, time\_on\_exercise\_ratio, detailed\_pages\_ratio, outline\_pages\_ratio, revision\_pages\_ratio, exercise\_ratio,

$((\text{time\_on\_detailed\_pages\_ratio}) * (\text{detailed\_pages\_ratio})) + ((1 - \text{exercise\_ratio}) * (\text{time\_on\_exercise\_ratio})) + ((\text{time\_on\_revision}) * (\text{revision\_pages\_ratio}))$  **as** sensing,

$((1 - \text{time\_on\_detailed\_pages\_ratio}) * (1 - \text{detailed\_pages\_ratio})) + ((\text{exercise\_ratio}) * (\text{time\_on\_exercise\_ratio})) + ((\text{time\_on\_revision}) * (\text{revision\_pages\_ratio}))$  **as** intuitive,

$((\text{time\_on\_detailed\_pages\_ratio}) * (\text{detailed\_pages\_ratio})) + ((1 - \text{exercise\_ratio}) * (\text{time\_on\_exercise\_ratio})) + ((\text{time\_on\_revision}) * (\text{revision\_pages\_ratio}))$   
+

$((1 - \text{time\_on\_detailed\_pages\_ratio}) * (1 - \text{detailed\_pages\_ratio})) + ((\text{exercise\_ratio}) * (\text{time\_on\_exercise\_ratio})) + ((\text{time\_on\_revision}) * (\text{revision\_pages\_ratio}))$  **as** total

**from** vw\_LS\_sensing\_intuitive\_patterns\_p

Estimating sequential global learning style

**select**

vw\_LS\_sequential\_global\_patterns\_p.user\_id,  
vw\_LS\_sequential\_global\_patterns\_p.regno,  
vw\_LS\_sequential\_global\_patterns\_p.global\_pages\_ratio,  
vw\_LS\_sequential\_global\_patterns\_p.sequential\_pages\_ratio,  
vw\_LS\_sequential\_global\_patterns\_p.outline\_pages\_ratio,  
vw\_LS\_sequential\_global\_patterns\_p.global\_pages\_time\_ratio,  
vw\_LS\_sequential\_global\_patterns\_p.sequential\_pages\_time\_ratio,  
vw\_LS\_sequential\_global\_patterns\_p.outline\_pages\_time\_ratio,  
vw\_LS\_sequential\_global\_patterns\_p.exercise\_score,

$(\text{global\_pages\_ratio} * \text{global\_pages\_time\_ratio}) * (\text{outline\_pages\_ratio} * \text{outline\_pages\_time\_ratio}) + ((1 - \text{sequential\_pages\_ratio}) * (1 - \text{sequential\_pages\_time\_ratio}))$  **as** global\_p,

```
(sequential_pages_ratio*sequential_pages_time_ratio)*
((1-outline_pages_ratio)*(1-outline_pages_time_ratio)) +
((1-global_pages_ratio)*(1-global_pages_time_ratio)) as sequential,
```

```
(global_pages_ratio*global_pages_time_ratio)*(outline_pages_ratio*outline_pages_time_ratio)+
((1-sequential_pages_ratio)*(1-sequential_pages_time_ratio))
```

```
+
(sequential_pages_ratio*sequential_pages_time_ratio)*
((1-outline_pages_ratio)*(1-outline_pages_time_ratio)) +
((1-global_pages_ratio)*(1-global_pages_time_ratio)) as total
```

```
from vw_LS_sequential_global_patterns_p
```

Estimating visual verbal learning style

```
select user_id,regno,
verbal_pages_ratio,
visual_pages_ratio,
verbal_time_ratio,
visual_time_ratio,
```

```
(verbal_pages_ratio*verbal_time_ratio)+((1-visual_pages_ratio)*(1-visual_time_ratio)) as p_verbal,
(visual_pages_ratio*visual_time_ratio)+((1-verbal_pages_ratio)*(1-verbal_time_ratio)) as
p_visual,(verbal_pages_ratio*verbal_time_ratio)+((1-visual_pages_ratio)*(1-visual_time_ratio)) +
(visual_pages_ratio*visual_time_ratio)+((1-verbal_pages_ratio)*(1-verbal_time_ratio)) as p_total from
vw_LS_verbal_visual_patterns_p
```

### Sample code matching contents with LS and CT

Learning materials are displayed to each student, according to their learning behaviors. These behaviors are paired and have values between 0 and 1. For non-zero value behaviors, the behavior with the larger value is selected in each pair. We then find content that matches these behaviors. Some students have no behaviors defined in the system. For these students, we find their content according to the cluster to which they belong. To do this, we find other students in the same cluster who have their behaviors in the system. We use these behaviors to find matching content. Find the learning characteristics of the logged-in user. If characteristics are not defined, we use those for cluster mates.

```
/**Get learning behaviors
 * @param int $studentId
 * @return array learner's behaviors
 */
private function getStudentCharacteristics(int $studentId): array
{
    return StudentCharacteristic::find()->select(['characteristicId', 'value'])
        ->where(['studentId' => $studentId])
        ->asArray()->all();
}
```

If a student has no characteristics and doesn't belong in any cluster, we redirect to an alert page.

```
if(empty($studentCharacteristics)){
```

```

        // If a student has no characteristic and doesn't belong in any cluster, we redirect to an alert page
        if(is_null(Yii::$app->user->identity->cluster)){
            return $this->redirect(['/reading/alert',
                'message' => 'We are unable to find your learning materials because you do not
belong in any cluster.']);
        } }

```

If a student has no behaviors, but they belong to a cluster, we use the behaviors of their cluster mates.

```
/**
```

```

 * Since we can't have all similar characteristics for all members in a cluster, we find the characteristics of all cluster
members, combine them and remove duplicates. */

```

```
$users = User::find()->select(['id']->where(['cluster' => Yii::$app->user->identity->cluster])
```

```

->andWhere(['not', ['id' => Yii::$app->user->identity->id]])

```

```

->asArray()->all();

```

```
$clusterMatesIds = [];
```

```
$clusterMatesCharacteristics = [];
```

```
foreach ($users as $user){
```

```

    $clusterMatesIds[] = $this->getStudentCharacteristics($user['id']);

```

```
}
```

```
foreach ($clusterMatesIds as $clusterMateIds){
```

```

    foreach ($clusterMateIds as $characteristic){

```

```

        $clusterMatesCharacteristics[] = $characteristic; } }

```

```
$studentCharacteristics = array_unique($clusterMatesCharacteristics, SORT_REGULAR);
```

```
/** After we find all the characteristics of a student or those of their cluster mates, we want to group them in their
pairs. */
```

```
$morphedCharacteristics = [];
```

```
foreach ($studentCharacteristics as $studentCharacteristic){
```

```

    $characteristicId = $studentCharacteristic['characteristic'];

```

```

    $morphedCharacteristics[$characteristicId] = (double)$studentCharacteristic['value'];

```

```
}
```

```
$pairsWithDuplicates = [];
```

```
foreach ($morphedCharacteristics as $key => $morphedCharacteristic){
```

```

// Find the pair to this student's characteristic
$pairedCharacteristic = Characteristic::find()->select(['pairedWith'])
    ->where(['id' => $key])->asArray()->one();

$pairedCharacteristicId = $pairedCharacteristic['pairedWith'];

$pairsWithDuplicates[] = [
    $key => $morphedCharacteristic,
    $pairedCharacteristicId => $morphedCharacteristics[$pairedCharacteristicId] ];
}

/** In each pair, we want to select a behavior with the largest value. We disregard behaviors with zero values */
$preferredCharacteristics = []; // We will match content according to these behaviors.

$pairs = array_unique($pairsWithDuplicates, SORT_REGULAR);

foreach ($pairs as $pair){
    $keys = [];
    $idx = 0;
    foreach ($pair as $pairKey => $pairValue){
        $keys[$idx] = $pairKey;
        $idx++;
    }
    if($pair[$keys[0]] > $pair[$keys[1]]){
        $preferredCharacteristics[] = $keys[0];
    }elseif($pair[$keys[0]] < $pair[$keys[1]]){
        $preferredCharacteristics[] = $keys[1];
    }else{
        if($pair[$keys[0]] > 0){
            $preferredCharacteristics[] = $keys[0];
        }
    }
}

```

We then read content that matches the “preferredCharacteristics” of a student

```
// Read content matching the student's behaviors
```

```

$materialsearchModel = new ReadingMaterialSearch();

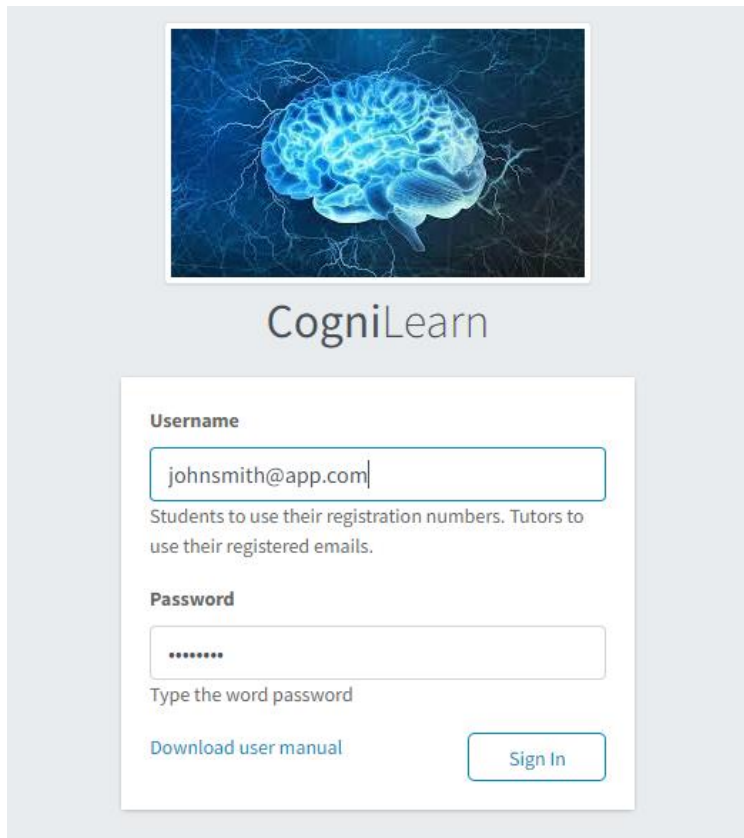
$materialDataProvider = $materialsearchModel->search(Yii::$app->request->queryParams, [
    'id' => Yii::$app->user->identity->id,
    'types' => $preferredCharacteristics
]);

$query = Content::find()->alias('cont')->select([
    'cont.id',
    'cont.courseId',
    'cont.url',
    'cont.description',
    'cont.createdBy',
    'cont.createdAt',
    'cont.type',
    'cont.topic'
])
->joinWith(['course cs' => function(ActiveQuery $q){
    $q->select(['cs.id', 'cs.code', 'cs.name']);
}], true, 'INNER JOIN')
->joinWith(['creator cr' => function(ActiveQuery $q){
    $q->select(['cr.id', 'cr.username']);
}], true, 'INNER JOIN')
->joinWith(['characteristic ch' => function(ActiveQuery $q){
    $q->select(['ch.id', 'ch.name', 'ch.description', 'ch.level']);
}], true, 'INNER JOIN')
->where(['in', 'cont.type', $additionalParams['types']])
->orderBy(['ch.level' => SORT_ASC, 'cont.id' => SORT_DESC])->toArray();

```

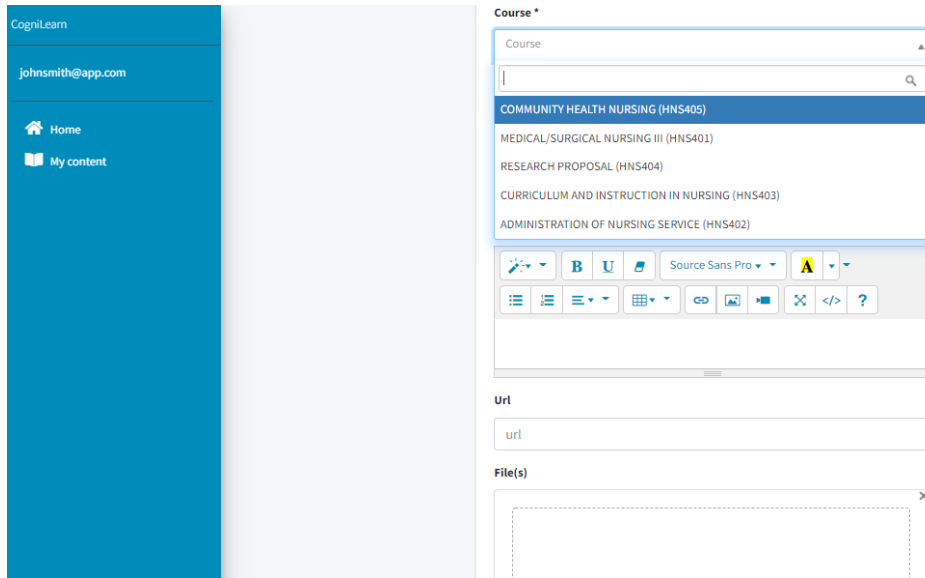
*Appendix 14. Model Interface and Functionalities.*

A tutor can logon <http://cognilearn.uonbi.ac.ke/> and then enter the assigned username and password

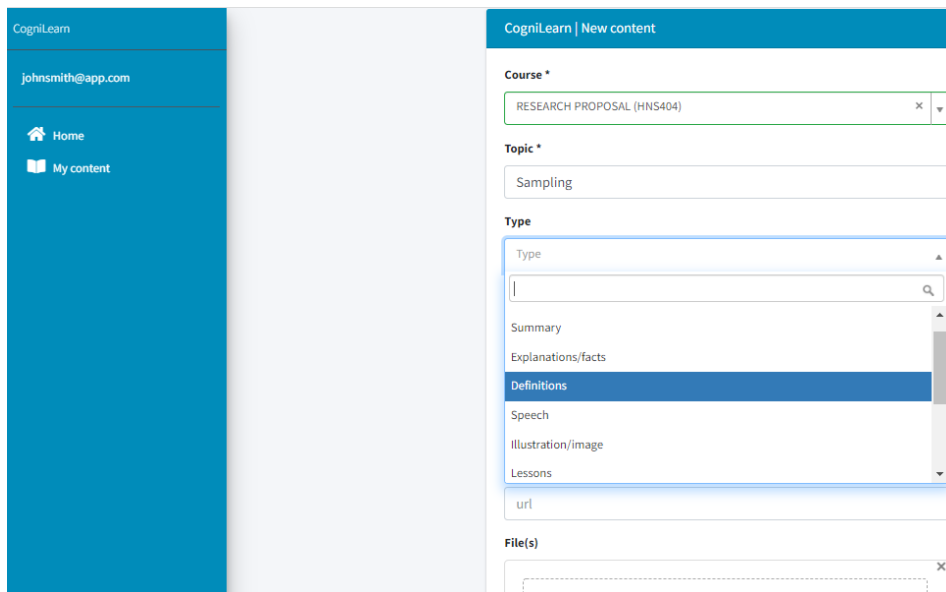


The image shows the CogniLearn login interface. At the top, there is a glowing blue brain icon. Below it, the text "CogniLearn" is displayed. The login form contains a "Username" field with the text "johnsmith@app.com" and a note: "Students to use their registration numbers. Tutors to use their registered emails." Below that is a "Password" field with masked characters "\*\*\*\*\*" and the instruction "Type the word password". At the bottom left, there is a link "Download user manual" and a "Sign In" button.

a tutor can select a course and add contents

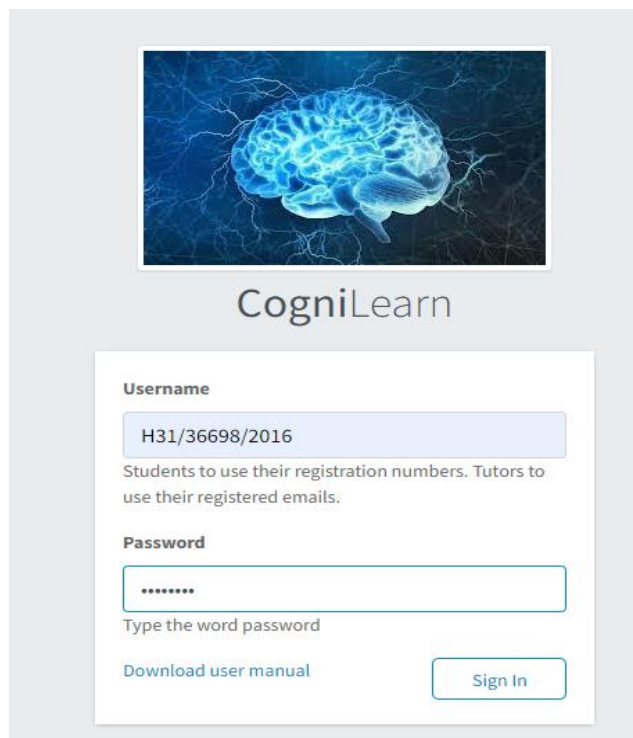


A tutor can specify the content type from the drop-down list. The content type matches individual learning behavior.



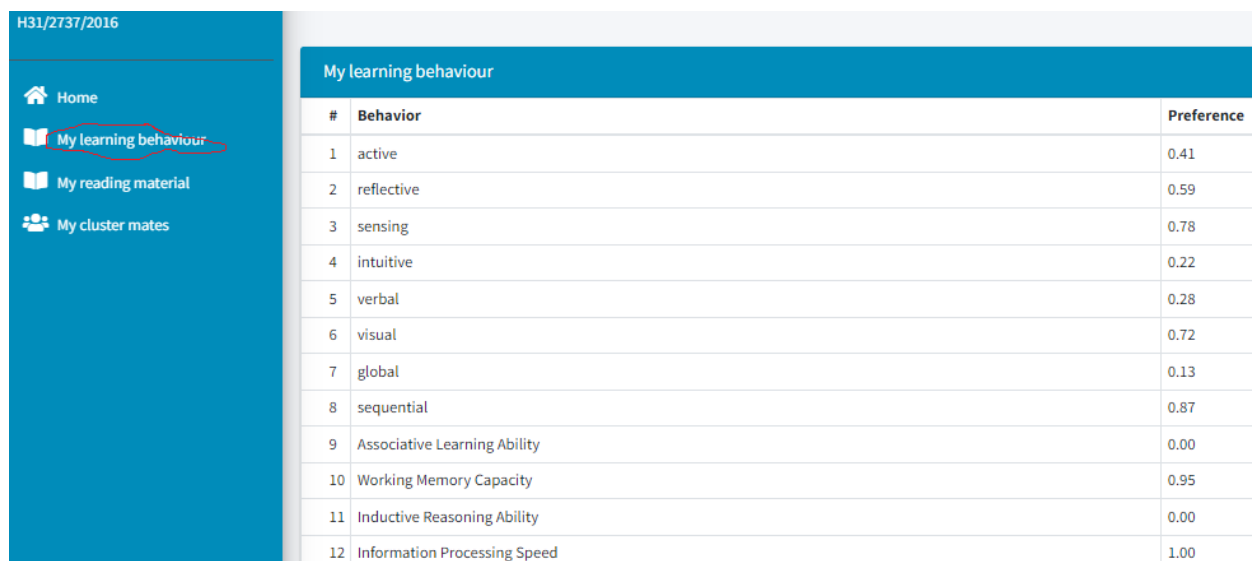


A student can log in using an assigned username and password.



The image shows the CogniLearn login interface. At the top, there is a glowing blue brain icon. Below it, the text "CogniLearn" is displayed. The login form contains two input fields: "Username" with the value "H31/36698/2016" and "Password" with masked characters "\*\*\*\*\*". A note below the username field states: "Students to use their registration numbers. Tutors to use their registered emails." Below the password field, it says "Type the word password". At the bottom, there are two buttons: "Download user manual" and "Sign In".

A student can view learning behavior on the dashboard.



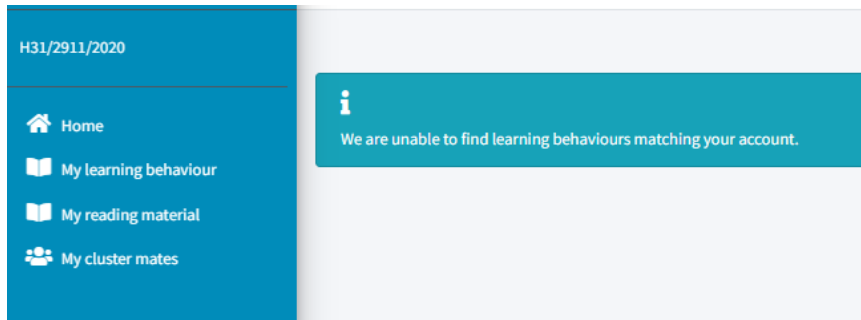
The image shows a dashboard for a student with the username "H31/2737/2016". The left sidebar contains navigation links: "Home", "My learning behaviour" (highlighted with a red circle), "My reading material", and "My cluster mates". The main content area is titled "My learning behaviour" and displays a table with 12 rows of learning behaviors and their preference scores.

#	Behavior	Preference
1	active	0.41
2	reflective	0.59
3	sensing	0.78
4	intuitive	0.22
5	verbal	0.28
6	visual	0.72
7	global	0.13
8	sequential	0.87
9	Associative Learning Ability	0.00
10	Working Memory Capacity	0.95
11	Inductive Reasoning Ability	0.00
12	Information Processing Speed	1.00

A student can also view learning materials matching the behavior.

#	+	Topic	Course code	Course name	Url	Created by	Created at
1	+	Introduction	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00
2	+	Osteoporosis	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00
3	+	Nursing Care Planning and Goals	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00
4	+	Anatomy of the Skeletal System	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00

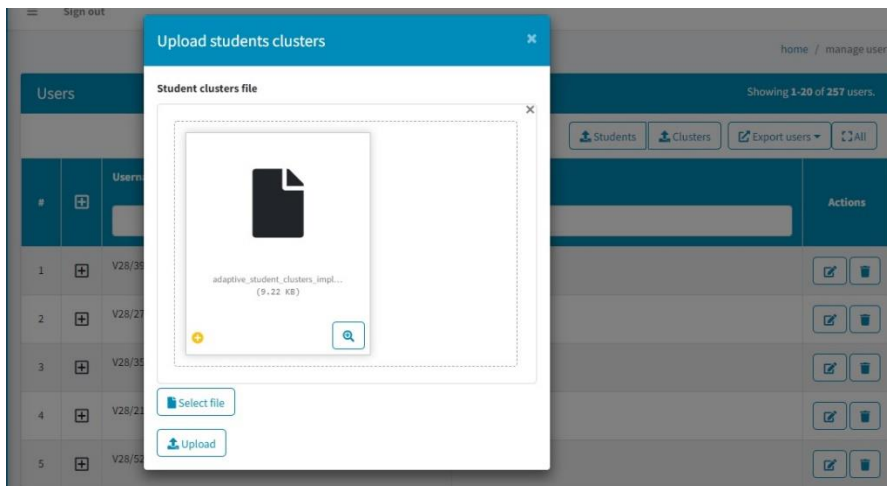
When the only available information is the cluster a student belongs to but learning behavior unknown.



Using this feature, a course lecturer can upload a file with learner profiles.

Users				Showing 1-20 of 257 users.
				<a href="#">Students</a> <a href="#">Clusters</a> <a href="#">Export users</a> <a href="#">All</a>
#	+	Username	Cluster	Actions
1	+	V28/3975/2009	3	<a href="#">Edit</a> <a href="#">Delete</a>
2	+	V28/2725/2009	1	<a href="#">Edit</a> <a href="#">Delete</a>
3	+	V28/35111/2010	1	<a href="#">Edit</a> <a href="#">Delete</a>
4	+	V28/2165/2012	2	<a href="#">Edit</a> <a href="#">Delete</a>
5	+	V28/52343/2012	3	<a href="#">Edit</a> <a href="#">Delete</a>

### File uploading feature of cognilearn



A student can also view cluster members with closer learning behavior. The system allocates similar learning contents to those with closer behavior

H31/2911/2020

- Home
- My learning behaviour
- My reading material
- My cluster mates

### Cluster 1

#	Username
1	V28/2725/2009
2	V28/35111/2010
3	H31/137062/2020

The system checks and allocates the new student contents similar to those in the same clusters.

H31/2737/2016

- Home
- My learning behaviour
- My reading material
- My cluster mates

### Content

Showing 1-5 of 5 content.

Export content All

#		Topic	Course code	Course name	Url	Created by	Created at
1	+	Introduction	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00
2	+	Osteoporosis	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00
3	+	Nursing Care Planning and Goals	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00
4	+	Anatomy of the Skeletal System	HNS401	MEDICAL/SURGICAL NURSING III		admin@app.com	04/01/22 16:00

*Appendix 15. Index of Learning Styles (ILS) Questionnaire and the scoresheet.*

**Index of Learning Styles (ILS)  
Learning Style Questionnaire**

*This questionnaire is designed to find out what your learning preferences are. It was originally designed by Felder and Silverman at North Carolina State University, USA.*

**Directions**

To complete the questionnaire please circle "a" or "b" to indicate your answer to every question. You may only choose one answer for each question and you must answer every question. If both "a" and "b" seem to apply to you, please choose the one that applies more frequently.

1. I understand something better after I
  - (a) try it out.
  - (b) think it through.
  
2. I would rather be considered
  - (a) realistic.
  - (b) innovative.
  
43. I tend to picture places I have been
  - (a) easily and fairly accurately.
  - (b) with difficulty and without much detail.
  
44. When solving problems in a group, I would be more likely to
  - (a) think of the steps in the solution process.
  - (b) think of possible consequences or applications of the solution in a wide range of areas.

Now turn to the scoring sheet so you can see where your preferences are.

## Learning Styles Questionnaire Scoring Sheet

1. Place a "1" in the appropriate spaces in the table below (e.g. if you answered "a" to Question 3, put a "1" in Column "a" by Question 3).
2. Add up the columns and write the totals in the indicated spaces.
3. For each of the four scales, subtract the smaller total from the larger one. Write the difference (1 to 11) and the letter (a or b) with the larger total.

Activist/Reflector			Sensing/Intuitive			Visual/Verbal			Sequential/Global		
Q	a	b	Q	a	b	Q	a	b	Q	a	b
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		
41			42			43			44		
<i>Total (add up each column)</i>											
Activist/Reflector			Sensing/Intuitive			Visual/Verbal			Sequential/Global		
Q	a	b	Q	a	b	Q	a	b	Q	a	b
<i>Larger – Smaller + Letter of Larger (see below*)</i>											

*\*Example: If your total was 3 for a and 8 for b: 8 – 3 = 5, b is letter of larger so you would enter 5b.*

### Explanation of scores

- If your score on a scale is 1-3, you have a mild preference for one or the other dimension but you are essentially well-balanced.
- If your score on a scale is 5-7, you have a moderate preference for one dimension of the scale and will learn more easily in a teaching environment that favors that dimension.
- If your score on a scale is 9-11, you have a strong preference for one dimension of the scale. You may have real difficulty learning in an environment that does not support that preference.

*Appendix 16. Survey data collected from ILS questionnaire.*

According to [cite], results in 1-3 are classified as mild, 3-5 Moderate, and 9-11 strong. In the current study, 0 means no preference, 1 low (mild), 2 moderate, and 3 high preferences.

student	active	reflective	sensing	intuitive	visual	verbal	sequential	global
1.	1	0	2	0	2	0	2	0
2.	0	2	1	0	0	1	2	0
3.	0	1	1	0	2	0	1	0
4.	0	2	2	0	0	1	1	0
5.	0	2	1	0	2	0	1	0
6.	1	0	2	0	0	1	0	2
7.	2	0	3	0	1	0	2	0
8.	0	1	1	0	3	0	0	1
9.	2	0	1	0	3	0	0	2
10.	0	1	0	2	2	0	1	0
11.	1	0	0	1	3	0	0	1
12.	0	1	1	0	2	0	2	0
13.	1	0	2	0	1	0	3	0
14.	0	1	2	0	0	1	1	0
15.	0	1	1	0	1	0	0	1
16.	0	1	2	0	1	0	0	1
17.	0	1	2	0	2	0	1	0
18.	0	2	0	3	3	0	0	1
19.	0	1	2	0	1	0	2	0
20.	2	0	2	0	1	0	0	1
21.	2	0	0	1	2	0	0	1
22.	0	2	1	0	1	0	1	0
23.	1	0	2	0	2	0	2	0
24.	0	2	0	1	2	0	1	0
25.	0	1	1	0	1	0	2	0
26.	1	0	3	0	1	0	3	0
27.	0	1	0	1	3	0	0	2
28.	1	0	0	1	1	0	1	0
29.	1	0	1	0	2	0	1	0
30.	0	1	1	0	3	0	1	0
31.	0	2	2	0	2	0	0	2
32.	2	0	2	0	3	0	2	0
33.	1	0	3	0	0	1	3	0
34.	0	2	0	1	3	0	1	0
35.	0	1	1	0	1	0	1	0
36.	0	2	0	2	1	0	1	0
37.	0	3	2	0	0	2	2	0
38.	0	3	2	0	2	0	3	0
39.	1	0	0	2	1	0	0	1

student	active	reflective	sensing	intuitive	visual	verbal	sequential	global
40.	0	2	1	0	0	1	1	0
41.	1	0	2	0	2	0	2	0
42.	3	0	0	2	2	0	0	2
43.	2	0	0	2	2	0	2	0
44.	1	0	1	0	1	0	2	0
45.	1	0	1	0	3	0	1	0
46.	2	0	0	2	3	0	0	1
47.	1	0	2	0	0	1	0	2
48.	1	0	1	0	0	1	1	0
49.	1	0	2	0	1	0	1	0
50.	2	0	0	1	3	0	1	0
51.	0	2	0	3	2	0	0	3
52.	0	1	2	0	3	0	0	2
53.	0	2	2	0	2	0	0	1
54.	1	0	1	0	2	0	3	0
55.	1	0	1	0	2	0	1	0
56.	1	0	0	1	1	0	2	0
57.	1	0	2	0	2	0	0	1
58.	1	0	2	0	3	0	2	0
59.	0	1	2	0	1	0	0	1
60.	0	2	0	3	2	0	0	0
61.	2	0	2	0	1	0	2	0
62.	2	0	0	1	2	0	0	2
63.	0	1	2	0	2	0	3	0
64.	2	0	1	0	2	0	0	1
65.	2	0	0	2	0	3	2	0
66.	2	0	3	0	3	0	2	0
67.	2	0	0	1	3	0	1	0
68.	1	0	0	1	3	0	1	0
69.	0	1	2	0	1	0	2	0
70.	0	2	1	0	2	0	1	0
71.	0	2	1	0	3	0	1	0
72.	0	2	0	1	1	0	1	0
73.	1	0	1	0	2	0	1	0
74.	0	1	1	0	0	1	0	1
75.	1	0	1	0	1	0	0	2
76.	2	0	1	0	0	2	2	0
77.	0	1	0	1	2	0	1	0
78.	0	1	2	0	0	3	0	1
79.	0	3	2	0	0	3	1	0
80.	0	1	2	0	2	0	2	0
81.	1	0	0	1	3	0	0	1
82.	1	0	0	1	2	0	0	2



student	active	reflective	sensing	intuitive	visual	verbal	sequential	global
83.	0	1	0	3	2	0	1	0
84.	0	1	0	1	1	0	0	1
85.	0	1	1	0	1	0	1	0
86.	0	1	1	0	3	0	0	1
87.	0	2	2	0	1	0	1	0
88.	1	0	0	1	0	1	2	0
89.	1	0	1	0	2	0	2	0
90.	0	1	2	0	0	1	1	0
91.	1	0	1	0	1	0	1	0
92.	1	0	2	0	2	0	2	0
93.	0	2	1	0	3	0	1	0
94.	0	1	0	3	2	0	1	0
95.	1	0	1	0	2	0	3	0
96.	0	1	2	0	2	0	1	0
97.	1	0	0	1	2	0	2	0
98.	0	0	0	0	0	0	0	0
99.	0	2	2	0	0	1	1	0
100.	1	0	0	1	1	0	0	1
101.	0	1	3	0	2	0	2	0
102.	0	1	0	2	1	0	0	1
103.	1	0	0	1	3	0	2	0
104.	0	1	1	0	1	0	0	1
105.	1	0	1	0	2	0	1	0
106.	0	1	1	0	1	0	0	1
107.	2	0	0	2	1	0	0	2
108.	0	1	0	1	0	2	0	2
109.	0	1	0	1	3	0	0	1
110.	0	1	0	1	2	0	1	0
111.	1	0	2	0	2	0	0	1
112.	1	0	1	0	1	0	0	2
113.	1	0	2	0	0	1	2	0
114.	0	2	1	0	2	0	2	0
115.	0	1	2	0	1	0	1	0
116.	0	1	0	3	2	0	1	0
117.	1	0	0	1	0	2	0	2
118.	1	0	2	0	2	0	0	1
119.	0	1	1	0	2	0	1	0
120.	2	0	0	2	0	1	0	1
121.	0	1	2	0	0	1	1	0
122.	0	2	3	0	0	1	2	0
123.	1	0	1	0	2	0	1	0
124.	0	2	0	3	0	3	0	3

Appendix 17. Classifications of survey results collected from ILS questionnaire.

Student	active - reflective	sensing-intuitive	visual-verbal	sequential-global
1.	active	sensing	visual	Sequential
2.	reflective	sensing	verbal	Sequential
3.	reflective	sensing	visual	Sequential
4.	reflective	sensing	verbal	Sequential
5.	reflective	sensing	visual	Sequential
6.	active	sensing	verbal	global
7.	active	sensing	visual	Sequential
8.	reflective	sensing	visual	global
9.	active	sensing	visual	global
10.	reflective	intuitive	visual	Sequential
11.	active	intuitive	visual	global
12.	reflective	sensing	visual	Sequential
13.	active	sensing	visual	Sequential
14.	reflective	sensing	verbal	Sequential
15.	reflective	sensing	visual	global
16.	reflective	sensing	visual	global
17.	reflective	sensing	visual	Sequential
18.	reflective	intuitive	visual	global
19.	reflective	sensing	visual	Sequential
20.	active	sensing	visual	global
21.	active	intuitive	visual	global
22.	reflective	sensing	visual	Sequential
23.	active	sensing	visual	Sequential
24.	reflective	intuitive	visual	Sequential
25.	reflective	sensing	visual	Sequential
26.	active	sensing	visual	Sequential
27.	reflective	intuitive	visual	global
28.	active	intuitive	visual	Sequential
29.	active	sensing	visual	Sequential
30.	reflective	sensing	visual	Sequential
31.	reflective	sensing	visual	global
32.	active	sensing	visual	Sequential
33.	active	sensing	verbal	Sequential
34.	reflective	intuitive	visual	Sequential
35.	reflective	sensing	visual	Sequential
36.	reflective	intuitive	visual	Sequential
37.	reflective	sensing	verbal	Sequential
38.	reflective	sensing	visual	Sequential
39.	active	intuitive	visual	global
40.	reflective	sensing	verbal	Sequential
41.	active	sensing	visual	Sequential

<b>Student</b>	<b>active - reflective</b>	<b>sensing-intuitive</b>	<b>visual-verbal</b>	<b>sequential-global</b>
42.	active	intuitive	visual	global
43.	active	intuitive	visual	Sequential
44.	active	sensing	visual	Sequential
45.	active	sensing	visual	Sequential
46.	active	intuitive	visual	global
47.	active	sensing	verbal	global
48.	active	sensing	verbal	Sequential
49.	active	sensing	visual	Sequential
50.	active	intuitive	visual	Sequential
51.	reflective	intuitive	visual	global
52.	reflective	sensing	visual	global
53.	reflective	sensing	visual	global
54.	active	sensing	visual	Sequential
55.	active	sensing	visual	Sequential
56.	active	intuitive	visual	Sequential
57.	active	sensing	visual	global
58.	active	sensing	visual	Sequential
59.	reflective	sensing	visual	global
60.	reflective	intuitive	visual	global
61.	active	sensing	visual	Sequential
62.	active	intuitive	visual	global
63.	reflective	sensing	visual	Sequential
64.	active	sensing	visual	global
65.	active	intuitive	verbal	Sequential
66.	active	sensing	visual	Sequential
67.	active	intuitive	visual	Sequential
68.	active	intuitive	visual	Sequential
69.	reflective	sensing	visual	Sequential
70.	reflective	sensing	visual	Sequential
71.	reflective	sensing	visual	Sequential
72.	reflective	intuitive	visual	Sequential
73.	active	sensing	visual	Sequential
74.	reflective	sensing	verbal	global
75.	active	sensing	visual	global
76.	active	sensing	verbal	Sequential
77.	reflective	intuitive	visual	Sequential
78.	reflective	sensing	verbal	global
79.	reflective	sensing	verbal	Sequential
80.	reflective	sensing	visual	Sequential
81.	active	intuitive	visual	global
82.	active	intuitive	visual	global
83.	reflective	intuitive	visual	Sequential

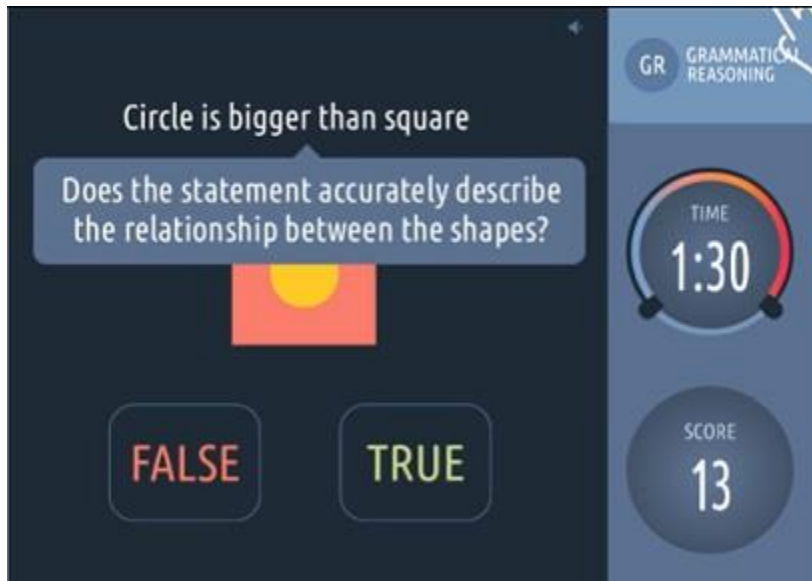
<b>Student</b>	<b>active - reflective</b>	<b>sensing-intuitive</b>	<b>visual-verbal</b>	<b>sequential-global</b>	
84.	reflective	intuitive	visual	global	
85.	reflective	sensing	visual	Sequential	
86.	reflective	sensing	visual	global	
87.	reflective	sensing	visual	Sequential	
88.	active	intuitive	verbal	Sequential	
89.	active	sensing	visual	Sequential	
90.	reflective	sensing	verbal	Sequential	
91.	active	sensing	visual	Sequential	
92.	active	sensing	visual	Sequential	
93.	reflective	sensing	visual	Sequential	
94.	reflective	intuitive	visual	Sequential	
95.	active	sensing	visual	Sequential	
96.	reflective	sensing	visual	Sequential	
97.	active	intuitive	visual	Sequential	
98.	reflective	intuitive	verbal	global	
99.	reflective	sensing	verbal	Sequential	
100.	active	intuitive	visual	global	
101.	reflective	sensing	visual	Sequential	
102.	reflective	intuitive	visual	global	
103.	active	intuitive	visual	Sequential	
104.	reflective	sensing	visual	global	
105.	active	sensing	visual	Sequential	
106.	reflective	sensing	visual	global	
107.	active	intuitive	visual	global	
108.	reflective	intuitive	verbal	global	
109.	reflective	intuitive	visual	global	
110.	reflective	intuitive	visual	Sequential	
111.	active	sensing	visual	global	
112.	active	sensing	visual	global	
113.	active	sensing	verbal	Sequential	
114.	reflective	sensing	visual	Sequential	
115.	reflective	sensing	visual	Sequential	
116.	reflective	intuitive	visual	Sequential	
117.	active	intuitive	verbal	global	
118.	active	sensing	visual	global	
119.	reflective	sensing	visual	Sequential	
120.	active	intuitive	verbal	global	
121.	reflective	sensing	verbal	Sequential	
122.	reflective	sensing	verbal	Sequential	
123.	active	sensing	visual	Sequential	
124.	reflective	intuitive	verbal	global	

*Appendix 18. Cognitive Assessment Tests*

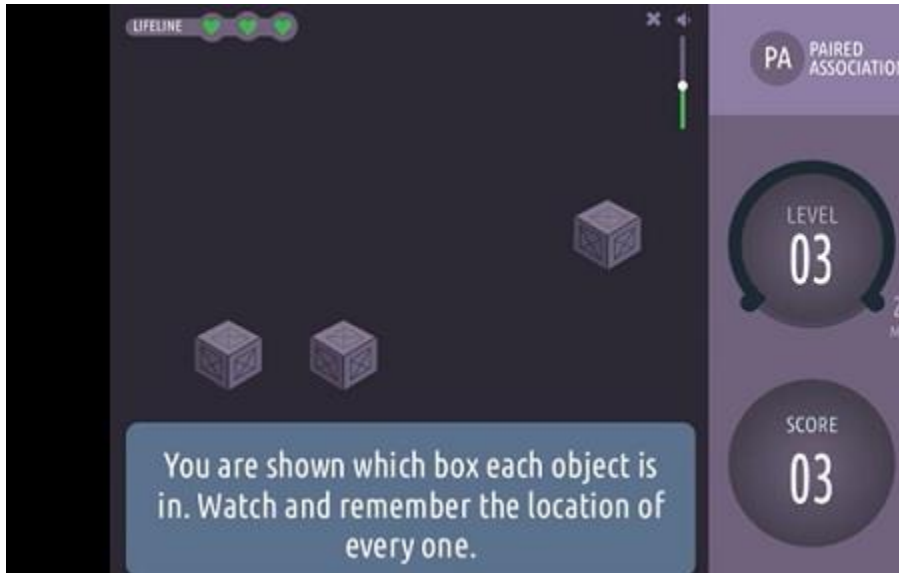
Span digit tests measure short-term memory i.e. ability to concentrate and remember (source: <https://www.cambridgebrainsciences.com/science/tasks>).



Grammatical reasoning tests the ability to quickly understand and make valid conclusions about concepts expressed in words.



Paired associate tests the ability to recall events, paired with the context in which they occurred.



Mental speed tests measure how quickly one can process information and make decisions based upon available facts i.e. the ability to come up with a quick solution ( source: <https://psychologytoday.tests.psychtests.com/>)

### Mental Speed Test

41 questions



### Mental Speed Test

41 questions



*Appendix 19. Classification of survey results for psychometric tests related to ALA.*

Any score below 7 out of 15 is low (weak), between 7 and 10 out of 15 is moderate, and above 10 out of 15 is high. The results were classified as 1 indicating low (mild), 2 moderate, and 3 high associative learning ability.

<b>Student</b>	<b>ALA score</b>	<b>Classification</b>
1	2	moderate
2	3	high
3	2	moderate
4	1	Low
5	3	high
6	3	high
7	2	moderate
8	1	Low
9	2	moderate
10	3	high
11	2	moderate
12	3	high
13	1	Low
14	2	moderate
15	3	high
16	2	moderate
17	3	high
18	3	high
19	1	Low
20	2	moderate
21	3	high
22	2	moderate
23	3	high
24	2	moderate
25	3	high
26	2	moderate
27	2	moderate
28	2	moderate
29	2	moderate
30	3	high
31	3	high
32	3	high
33	2	moderate
34	3	high
35	1	Low
36	1	Low
37	2	moderate

Student	ALA score	Classification
38	3	high
39	2	moderate
40	2	moderate
41	3	high
42	3	high
43	3	high
44	3	high
45	2	moderate
46	3	high
47	3	high
48	1	Low
49	3	high
50	2	moderate
51	2	moderate
52	2	moderate
53	3	high
54	2	moderate
55	2	moderate
56	1	Low
57	2	moderate
58	2	moderate
59	2	moderate
60	2	moderate
61	2	moderate
62	3	high
63	3	high
64	2	moderate
65	1	Low
66	3	high
67	3	high
68	3	high
69	1	Low
70	3	high
71	3	high
72	2	moderate
73	3	high
74	2	moderate
75	2	moderate
76	2	moderate
77	3	high
78	3	high
79	3	high



Student	ALA score	Classification
80	2	moderate
81	1	Low
82	3	high
83	3	high
84	1	Low
85	3	high
86	3	high
87	2	moderate
88	2	moderate
89	1	Low
90	2	moderate
91	2	moderate
92	3	high
93	1	Low
94	2	moderate
95	1	Low
96	2	moderate
97	2	moderate
98	2	moderate
99	2	moderate
100	3	high
101	2	moderate
102	2	moderate
103	3	high
104	3	high
105	3	high
106	3	high
107	1	Low
108	2	moderate
109	3	high
110	1	Low
111	2	moderate
112	2	moderate
113	2	moderate
114	2	moderate
115	3	high
116	3	high
117	2	moderate

*Appendix 20. Classification of survey results for psychometric tests related to IRA.*

Any score below 7 out of 15 was considered low (weak), between 7 and 10 out of 15 is moderate, and above 10 out of 15 is high. The results were classified as 1 indicating low (mild), 2 moderate, and 3 high inductive reasoning ability.

Student	IRA score	Classification
1	2	moderate
2	2	moderate
3	2	moderate
4	2	moderate
5	1	Low
6	1	Low
7	1	Low
8	2	moderate
9	1	Low
10	1	Low
11	2	moderate
12	2	moderate
13	3	High
14	3	High
15	1	Low
16	1	Low
17	3	High
18	1	Low
19	2	moderate
20	3	High
21	1	Low
22	2	moderate
23	1	Low
24	1	Low
25	1	Low
26	2	moderate
27	3	High
28	2	moderate
29	2	moderate
30	2	moderate
31	3	High
32	2	moderate
33	2	moderate
34	1	Low
35	2	moderate
36	2	moderate
37	3	High
38	1	Low
39	2	moderate
40	1	Low
41	3	High
42	3	High
43	2	moderate
44	3	High
45	2	moderate

Student	IRA score	Classification
46	2	moderate
47	2	moderate
48	3	High
49	2	moderate
50	2	moderate
51	2	moderate
52	3	High
53	3	High
54	2	moderate
55	3	High
56	3	High
57	1	Low
58	1	Low
59	1	Low
60	2	moderate
61	2	moderate
62	1	Low
63	2	moderate
64	2	moderate
65	1	Low
66	3	High
67	2	moderate
68	2	moderate
69	3	High
70	1	Low
71	1	Low
72	1	Low
73	1	Low
74	1	Low
75	1	Low
76	2	moderate
77	2	moderate
78	2	moderate
79	2	moderate
80	2	moderate
81	1	Low
82	1	Low
83	2	moderate
84	1	Low
85	3	High
86	1	Low
87	3	High
88	1	Low
89	2	moderate
90	2	moderate
91	2	moderate
92	2	moderate
93	2	moderate
94	2	moderate

Student	IRA score	Classification
95	1	Low
96	3	High
97	2	moderate
98	2	moderate
99	1	Low
100	2	Moderate
101	3	High
102	2	Moderate
103	2	Moderate
104	1	Low
105	3	High
106	1	Low
107	1	Low
108	2	Moderate
109	1	Low
110	1	Low
111	2	Moderate
112	1	Low
113	3	High

*Appendix 21. Classification of survey results for psychometric tests related to IPS.*

Any score below 7 out of 15 is low (weak), between 7 and 10 out of 15 is moderate, and above 10 out of 15 is high. The results were classified as 1 indicating low (mild), 2 moderate, and 3 high information processing speed.

Student	IPS score	Classification
1	3	High
2	2	Moderate
3	2	Moderate
4	2	Moderate
5	2	Moderate
6	2	Moderate
7	1	Low
8	1	Low
9	2	Moderate
10	2	Moderate
11	2	Moderate
12	2	Moderate
13	2	Moderate
14	3	High
15	2	Moderate
16	3	High
17	3	High
18	2	Moderate

Student	IPS score	Classification
19	2	Moderate
20	2	Moderate
21	3	High
22	1	Low
23	1	Low
24	2	Moderate
25	2	Moderate
26	2	Moderate
27	2	Moderate
28	2	Moderate
29	3	High
30	2	Moderate
31	3	High
32	1	Low
33	3	High
34	1	Low
35	1	Low
36	2	Moderate
37	2	Moderate
38	2	Moderate
39	2	Moderate
40	1	Low
41	3	High
42	1	Low
43	1	Low
44	1	Low
45	2	Moderate
46	1	Low
47	1	Low
48	1	Low
49	2	Moderate
50	3	High
51	1	Low
52	2	Moderate
53	1	Low
54	3	High
55	1	Low
56	3	High
57	2	Moderate
58	3	High
59	1	Low
60	1	Low

Student	IPS score	Classification
61	1	Low
62	2	Moderate
63	2	Moderate
64	3	High
65	1	Low
66	2	Moderate
67	2	Moderate
68	2	Moderate
69	2	Moderate
70	1	Low
71	2	Moderate
72	1	Low
73	2	Moderate
74	3	High
75	2	Moderate
76	2	Moderate
77	3	High
78	2	Moderate
79	2	Moderate
80	2	Moderate
81	3	High
82	2	Moderate
83	2	Moderate
84	2	Moderate
85	3	High
86	1	Low
87	2	Moderate
88	2	Moderate
89	3	High
90	2	Moderate
91	3	High
92	2	Moderate
93	2	Moderate
94	1	Low
95	3	High
96	3	High
97	3	High
98	2	Moderate
99	2	Moderate
100	1	Low
101	1	Low
102	2	Moderate

Student	IPS score	Classification
103	3	High
104	2	Moderate
105	3	High
106	2	Moderate
107	2	Moderate
108	2	Moderate
109	2	Moderate
110	1	Low
111	2	Moderate
112	2	Moderate
113	1	Low
114	2	Moderate

*Appendix 22. Classification of survey results for psychometric tests related to WMC.*

Any score below 7 out of 15 is low (weak), between 7 and 10 out of 15 is moderate, and above 10 out of 15 is high. The results were classified as 1 indicating low (mild), 2 moderate, and 3 high working memory capacity.

student	WMC score	Classification
1	2	Moderate
2	2	Moderate
3	3	High
4	3	High
5	2	Moderate
6	2	Moderate
7	3	High
8	3	High
9	3	High
10	3	High
11	3	High
12	3	High
13	3	High
14	2	Moderate
15	2	Moderate
16	3	High
17	1	Low
18	2	Moderate
19	3	High
20	3	High
21	3	High
22	2	Moderate

student	WMC score	Classification
23	2	Moderate
24	3	High
25	3	High
26	3	High
27	3	High
28	3	High
29	3	High
30	3	High
31	3	High
32	2	Moderate
33	1	Low
34	3	High
35	3	High
36	3	High
37	3	High
38	2	Moderate
39	3	High
40	3	High
41	3	High
42	2	Moderate
43	3	High
44	3	High
45	1	Low
46	3	High
47	1	Low
48	2	Moderate
49	2	Moderate
50	3	High
51	2	Moderate
52	3	High
53	2	Moderate
54	3	High
55	2	Moderate
56	3	High
57	2	Moderate
58	3	High
59	2	Moderate
60	2	Moderate
61	1	Low
62	2	Moderate
63	2	Moderate
64	2	Moderate



<b>student</b>	<b>WMC score</b>	<b>Classification</b>
65	3	High
66	3	High
67	2	Moderate
68	3	High
69	2	Moderate
70	3	High
71	2	Moderate
72	2	Moderate
73	1	Low
74	2	Moderate
75	2	Moderate
76	3	High
77	3	High
78	3	High
79	3	High
80	1	Low
81	3	High
82	3	High
83	1	Low
84	3	High
85	3	High
86	3	High
87	2	Moderate
88	2	Moderate
89	3	High
90	3	High
91	2	Moderate
92	3	High
93	3	High
94	3	High
95	3	High
96	2	Moderate
97	1	Low
98	3	High
99	3	High
100	3	High
101	3	High
102	2	Moderate
103	3	High
104	3	High
105	2	Moderate
106	2	Moderate

student	WMC score	Classification
107	2	Moderate
108	3	High
109	3	High
110	3	High

*Appendix 23. Classification of results predicted for active reflective LS.*

Student	Active	Reflective	Classification
1	0.39792	0.60208	Reflective
2	0.896453	0.103547	Active
3	0.641596	0.358404	Active
4	0.697992	0.302008	Active
5	0.89596	0.10404	Active
6	0.738906	0.261094	Active
7	0.503655	0.496345	Active
8	0.385741	0.614259	Reflective
9	0.473426	0.526574	Reflective
10	0.470901	0.529099	Reflective
11	0.688257	0.311743	Active
12	0.569113	0.430887	Active
13	0.29263	0.70737	Reflective
14	0.423472	0.576528	Reflective
15	0.556938	0.443062	Active
16	0.598157	0.401843	Active
17	0.412816	0.587184	Reflective
18	0.277424	0.722576	Reflective
19	0.382099	0.617901	Reflective
20	0.563487	0.436513	Active
21	0.324336	0.675664	Reflective
22	0.630083	0.369917	Active
23	0.297953	0.702047	Reflective
24	0.47514	0.52486	Reflective
25	0.654666	0.345334	Active
26	0.417081	0.582919	Reflective
27	0.860772	0.139228	Active
28	0.489711	0.510289	Reflective
29	0.50207	0.49793	Active
30	0.565218	0.434782	Active
31	0.448133	0.551867	Reflective
32	0.460291	0.539709	Reflective
33	0.537303	0.462697	Active
34	0.446308	0.553692	Reflective
35	0.565267	0.434733	Active

<b>Student</b>	<b>Active</b>	<b>Reflective</b>	<b>Classification</b>
36	0.411013	0.588987	Reflective
37	0.534973	0.465027	Active
38	0.605631	0.394369	Active
39	0.299676	0.700324	Reflective
40	0.646385	0.353615	Active
41	0.527844	0.472156	Active
42	0.597111	0.402889	Active
43	0.503789	0.496211	Active
44	0.365334	0.634666	Reflective
45	0.670776	0.329224	Active
46	0.270321	0.729679	Reflective
47	0.498396	0.501604	Reflective
48	0.667848	0.332152	Active
49	0.527337	0.472663	Active
50	0.506205	0.493795	Active
51	0.43632	0.56368	Reflective
52	0.530599	0.469401	Active
53	0.681038	0.318962	Active
54	0.678238	0.321762	Active
55	0.625768	0.374232	Active
56	0.524965	0.475035	Active
57	0.612683	0.387317	Active
58	0.616248	0.383752	Active
59	0.511133	0.488867	Active
60	0.552986	0.447014	Active
61	0.705528	0.294472	Active
62	0.277567	0.722433	Reflective
63	0.368375	0.631625	Reflective
64	0.403613	0.596387	Reflective
65	0.449774	0.550226	Reflective
66	0.575486	0.424514	Active
67	0.661347	0.338653	Active
68	0.694805	0.305195	Active
69	0.53326	0.46674	Active
70	0.818952	0.181048	Active
71	0.531822	0.468178	Active
72	0.334894	0.665106	Reflective
73	0.718881	0.281119	Active
74	0.851848	0.148152	Active
75	0.697471	0.302529	Active
76	0.754008	0.245992	Active
77	0.454795	0.545205	Reflective

<b>Student</b>	<b>Active</b>	<b>Reflective</b>	<b>Classification</b>
78	0.793413	0.206587	Active
79	0.338764	0.661236	Reflective
80	0.767284	0.232716	Active
81	0.59385	0.40615	Active
82	0.35504	0.64496	Reflective
83	0.832103	0.167897	Active
84	0.463274	0.536726	Reflective
85	0.701879	0.298121	Active
86	0.375749	0.624251	Reflective
87	0.683383	0.316617	Active
88	0.51092	0.48908	Active
89	0.730833	0.269167	Active
90	0.264911	0.735089	Reflective
91	0.345563	0.654438	Reflective
92	0.865537	0.134463	Active
93	0.849289	0.150711	Active
94	0.584789	0.415211	Active
95	0.628148	0.371852	Active
96	0.833389	0.166611	Active
97	0.776208	0.223792	Active
98	0.512504	0.487496	Active
99	0.841899	0.158101	Active
100	0.815264	0.184736	Active
101	0.763551	0.236449	Active
102	0.354829	0.645171	Reflective
103	0.544432	0.455568	Active
104	0.562922	0.437078	Active
105	0.412458	0.587542	Reflective
106	0.755402	0.244598	Active
107	0.37549	0.62451	Reflective
108	0.493788	0.506212	Reflective
109	0.322451	0.677549	Reflective
110	0.431297	0.568703	Reflective
111	0.51489	0.48511	Active
112	0.504627	0.495373	Active
113	0.398908	0.601092	Reflective
114	0.450097	0.549903	Reflective
115	0.600088	0.399912	Active
116	0.462254	0.537746	Reflective

Appendix 24. classification of results predicted for sensing intuitive LS.

Student	Sensing	Intuitive	Classification
1	0.739972	0.260028	Sensing
2	0.592349	0.407651	Sensing
3	0.694839	0.305161	Sensing
4	0.613534	0.386466	Sensing
5	0.770947	0.229053	Sensing
6	0.686745	0.313255	Sensing
7	0.661675	0.338325	Sensing
8	0.670677	0.329323	Sensing
9	0.588075	0.411925	Sensing
10	0.747542	0.252458	Sensing
11	0.615613	0.384387	Sensing
12	0.658432	0.341568	Sensing
13	0.486655	0.513345	Intuitive
14	0.586389	0.413611	Sensing
15	0.625858	0.374142	Sensing
16	0.613551	0.386449	Sensing
17	0.556976	0.443024	Sensing
18	0.561757	0.438243	Sensing
19	0.576229	0.423771	Sensing
20	0.528569	0.471431	Sensing
21	0.632597	0.367403	Sensing
22	0.621174	0.378826	sensing
23	0.595934	0.404066	sensing
24	0.617202	0.382798	sensing
25	0.501484	0.498516	sensing
26	0.578712	0.421288	sensing
27	0.469145	0.530855	intuitive
28	0.555241	0.444759	sensing
29	0.564388	0.435612	sensing
30	0.483433	0.516567	intuitive
31	0.514823	0.485177	sensing
32	0.481333	0.518667	intuitive
33	0.656915	0.343085	sensing
34	0.39709	0.60291	intuitive
35	0.51944	0.48056	sensing
36	0.493313	0.506687	intuitive
37	0.607949	0.392051	sensing
38	0.508271	0.491729	sensing
39	0.550944	0.449056	sensing
40	0.541366	0.458634	sensing
41	0.588803	0.411197	sensing

Student	Sensing	Intuitive	Classification
42	0.484881	0.515119	intuitive
43	0.525634	0.474366	sensing
44	0.508669	0.491331	sensing
45	0.43019	0.56981	intuitive
46	0.483279	0.516721	intuitive
47	0.464153	0.535847	intuitive
48	0.430451	0.569549	intuitive
49	0.404183	0.595817	intuitive
50	0.467606	0.532394	intuitive
51	0.581492	0.418508	sensing
52	0.381729	0.618271	intuitive
53	0.511925	0.488075	sensing
54	0.379079	0.620921	intuitive
55	0.56309	0.43691	sensing
56	0.389498	0.610502	intuitive
57	0.386028	0.613972	intuitive
58	0.396486	0.603514	intuitive
59	0.256159	0.743841	intuitive
60	0.337123	0.662877	intuitive
61	0.541743	0.458257	sensing
62	0.343338	0.656662	intuitive
63	0.372338	0.627662	intuitive
64	0.454698	0.545302	intuitive
65	0.378024	0.621976	intuitive
66	0.340754	0.659246	intuitive
67	0.321758	0.678242	intuitive
68	0.399864	0.600136	intuitive
69	0.347088	0.652912	intuitive
70	0.368219	0.631781	intuitive
71	0.309011	0.690989	intuitive
72	0.537508	0.462492	sensing
73	0.324189	0.675811	intuitive
74	0.317677	0.682323	intuitive
75	0.328476	0.671524	intuitive
76	0.445818	0.554182	intuitive
77	0.358027	0.641973	intuitive
78	0.332162	0.667838	intuitive
79	0.233114	0.766886	intuitive
80	0.295901	0.704099	intuitive
81	0.307218	0.692782	intuitive
82	0.289471	0.710529	intuitive
83	0.347964	0.652036	intuitive

<b>Student</b>	<b>Sensing</b>	<b>Intuitive</b>	<b>Classification</b>
84	0.293604	0.706396	intuitive
85	0.421012	0.578988	intuitive
86	0.294726	0.705274	intuitive
87	0.316486	0.683514	intuitive
88	0.465394	0.534606	intuitive
89	0.251998	0.748002	intuitive
90	0.239348	0.760652	intuitive
91	0.286133	0.713867	intuitive
92	0.243246	0.756754	intuitive
93	0.235306	0.764694	intuitive
94	0.211801	0.788199	intuitive
95	0.461215	0.538785	intuitive
96	0.20083	0.79917	intuitive
97	0.222303	0.777697	intuitive
98	0.183928	0.816072	intuitive
99	0.224649	0.775351	intuitive
100	0.232706	0.767294	intuitive
101	0.277779	0.722221	intuitive
102	0.240697	0.759303	intuitive
103	0.256996	0.743004	intuitive
104	0.113796	0.886204	intuitive
105	0.173435	0.826565	intuitive
106	0.159621	0.840379	intuitive
107	0.143691	0.856309	intuitive
108	0.110598	0.889402	intuitive
109	0.085979	0.914021	intuitive
110	0.094008	0.905992	intuitive
111	0.236981	0.763019	intuitive
112	0.160888	0.839112	intuitive
113	0.079626	0.920374	intuitive
114	0.081978	0.918022	intuitive
115	0.095659	0.904341	intuitive
116	0.087039	0.912961	intuitive

*Appendix 25. classification of results predicted for the sequential global LS.*

<b>Student</b>	<b>Global</b>	<b>Sequential</b>	<b>Classification</b>
1	0.718166	0.281834155	Global
2	0.593693	0.406307437	Global
3	0.591637	0.40836347	Global
4	0.47275	0.527249633	Sequential
5	0.449664	0.550336135	Sequential
6	0.439136	0.560864114	Sequential
7	0.566485	0.433515	Global
8	0.426435	0.573565299	Sequential
9	0.586825	0.413175	Global
10	0.411144	0.588856392	Sequential
11	0.3996	0.600399586	Sequential
12	0.603162	0.396838	Global
13	0.384922	0.615078236	Sequential
14	0.643237	0.356763	Global
15	0.327433	0.6725668	Sequential
16	0.320154	0.679846479	Sequential
17	0.689966	0.310034	Global
18	0.30462	0.695379665	Sequential
19	0.296173	0.703826777	Sequential
20	0.263564	0.73643595	Sequential
21	0.737608	0.262392	Global
22	0.252973	0.747027011	Sequential
23	0.251543	0.748457101	Sequential
24	0.24537	0.754629605	Sequential
25	0.231304	0.768695996	Sequential
26	0.23104	0.768959741	Sequential
27	0.774529	0.225471	Global
28	0.224278	0.775721722	Sequential
29	0.219171	0.780829162	Sequential
30	0.788718	0.211282	Global
31	0.206108	0.793891911	Sequential
32	0.198791	0.801209371	Sequential
33	0.187517	0.812483319	Sequential
34	0.182003	0.817997328	Sequential
35	0.181491	0.81850939	Sequential
36	0.826259	0.173741	Global
37	0.173234	0.826766195	Sequential
38	0.170889	0.829110588	Sequential
39	0.164861	0.835139166	Sequential
40	0.835342	0.164658	Global
41	0.153142	0.846857834	Sequential



Student	Global	Sequential	Classification
42	0.146297	0.853703055	Sequential
43	0.856423	0.143577	Global
44	0.133461	0.866539087	Sequential
45	0.132727	0.867273394	Sequential
46	0.870227	0.129773	Global
47	0.127045	0.872955224	Sequential
48	0.873856	0.126144	Global
49	0.124671	0.875328977	Sequential
50	0.123436	0.876563598	Sequential
51	0.122686	0.877313867	Sequential
52	0.119154	0.880846023	Sequential
53	0.115184	0.88481634	Sequential
54	0.105	0.894999866	Sequential
55	0.102625	0.897375135	Sequential
56	0.101725	0.898275084	Sequential
57	0.099631	0.900369287	Sequential
58	0.097544	0.902455557	Sequential
59	0.09543	0.904569508	Sequential
60	0.093515	0.906484699	Sequential
61	0.091824	0.908175851	Sequential
62	0.088662	0.911337943	Sequential
63	0.086247	0.913753407	Sequential
64	0.085362	0.914638341	Sequential
65	0.082826	0.917173877	Sequential
66	0.081068	0.918931665	Sequential
67	0.078793	0.921207075	Sequential
68	0.071679	0.928321038	Sequential
69	0.06995	0.93004953	Sequential
70	0.069888	0.93011162	Sequential
71	0.069448	0.930551597	Sequential
72	0.069401	0.930599231	Sequential
73	0.069373	0.930627486	Sequential
74	0.066795	0.933204889	Sequential
75	0.065054	0.934946229	Sequential
76	0.064426	0.935574351	Sequential
77	0.063018	0.936982162	Sequential
78	0.062714	0.937285543	Sequential
79	0.061262	0.938738439	Sequential
80	0.061032	0.938968432	Sequential
81	0.058553	0.9414468	Sequential
82	0.056232	0.943767769	Sequential
83	0.056194	0.943806219	Sequential

Student	Global	Sequential	Classification
84	0.054399	0.945601051	Sequential
85	0.051132	0.948867976	Sequential
86	0.04722	0.952779629	Sequential
87	0.035846	0.964153529	Sequential
88	0.0354	0.964600449	Sequential
89	0.034789	0.965210915	Sequential
90	0.96535	0.03465	Global
91	0.966017	0.033983	Global
92	0.966031	0.033969	Global
93	0.033388	0.966611525	Sequential
94	0.033337	0.96666293	Sequential
95	0.032999	0.967001143	Sequential
96	0.969339	0.030661	Global
97	0.030287	0.969713418	Sequential
98	0.97211	0.02789	Global
99	0.027653	0.972347304	Sequential
100	0.027543	0.972456675	Sequential
101	0.026788	0.973211757	Sequential
102	0.974841	0.025159	Global
103	0.025072	0.974928424	Sequential
104	0.023954	0.976045573	Sequential
105	0.976046	0.023954	Global
106	0.977337	0.022663	Global
107	0.022539	0.977461465	Sequential
108	0.022043	0.977957068	Sequential
109	0.021842	0.978157689	Sequential
110	0.021558	0.978442368	Sequential
111	0.021288	0.97871185	Sequential
112	0.021251	0.978749465	Sequential
113	0.020673	0.97932688	Sequential
114	0.019629	0.980370922	Sequential
115	0.018623	0.981377135	Sequential
116	0.018572	0.981427551	Sequential
117	0.018156	0.981844139	Sequential
118	0.017833	0.982166789	Sequential
119	0.016974	0.983025983	Sequential
120	0.016523	0.983476911	Sequential
121	0.014663	0.985337285	Sequential
122	0.014328	0.985671724	Sequential
123	0.014213	0.985787491	Sequential
124	0.014111	0.985888904	Sequential
125	0.013359	0.986640506	Sequential

Student	Global	Sequential	Classification
126	0.01323	0.986770401	Sequential
127	0.013195	0.986805349	Sequential
128	0.013153	0.986847274	Sequential
129	0.013119	0.986881275	Sequential
130	0.012637	0.987362997	Sequential
131	0.012267	0.98773337	Sequential
132	0.012254	0.987746261	Sequential
133	0.012065	0.987934843	Sequential
134	0.011914	0.9880864	Sequential
135	0.011792	0.988208437	Sequential
136	0.011106	0.988893605	Sequential
137	0.010729	0.98927121	Sequential
138	0.009957	0.990043158	Sequential
139	0.007895	0.992105153	Sequential
140	0.007605	0.992395124	Sequential
141	0.00718	0.992820391	Sequential
142	0.007074	0.992926462	Sequential
143	0.006989	0.993010723	Sequential
144	0.006326	0.993673902	Sequential
145	0.006274	0.993726466	Sequential
146	0.006162	0.993837523	Sequential
147	0.005858	0.994141596	Sequential
148	0.005511	0.994488517	Sequential
149	0.0055	0.994499575	Sequential
150	0.004357	0.995642579	Sequential
151	0.00407	0.995929638	Sequential
152	0.002435	0.997565155	Sequential
153	0.000835	0.999164659	Sequential

*Appendix 26. Classification of results predicted for the visual verbal LS.*

Student	Verbal	Visual	Classification
1	0.4	0.6	visual
2	0.21	0.79	visual
3	0.31	0.69	visual
4	0.32	0.68	visual
5	0.23	0.77	visual
6	0.05	0.95	visual
7	0.28	0.72	visual
8	0.39	0.61	visual
9	0.33	0.67	visual
10	0.12	0.88	visual

<b>Student</b>	<b>Verbal</b>	<b>Visual</b>	<b>Classification</b>
11	0.21	0.79	visual
12	0.13	0.87	visual
13	0.33	0.67	visual
14	0.22	0.78	visual
15	0.41	0.59	visual
16	0.1	0.9	visual
17	0.29	0.71	visual
18	0.28	0.72	visual
19	0.37	0.63	visual
20	0.36	0.64	visual
21	0.17	0.83	visual
22	0.23	0.77	visual
23	0.08	0.92	visual
24	0.21	0.79	visual
25	0.48	0.52	visual
26	0.24	0.76	visual
27	0.24	0.76	visual

*Appendix 27. Classification of results predicted for the ALA.*

<b>student</b>	<b>ALA</b>	<b>Classification</b>
1	0.4	Low
2	0.3333	Low
3	0.3333	Low
4	0.4118	Moderate
5	0.3333	Low
6	0.3333	Low
7	0.3333	Low
8	0.3333	Low
9	0.3333	Low
10	0.3333	Low
11	0.4	Low
12	0.4	Low
13	0.3333	Low
14	0.3333	Low
15	0.3333	Low
16	0.4	Low
17	0.4118	Moderate
18	0.4826	Moderate
19	0.3333	Low
20	0.4	Low
21	0.4	Low

<b>student</b>	<b>ALA</b>	<b>Classification</b>
22	0.375	Low
23	0.3333	Low
24	0.4472	Moderate
25	0.3333	Low
26	0.3333	Low
27	0.375	Low
28	0.3333	Low
29	0.4783	Moderate
30	0.3636	Low
31	0.3333	Low
32	0.4286	Moderate
33	0.3636	Low
34	0.4237	Moderate
35	0.3889	Low
36	0.4	Low
37	0.4211	Moderate
38	0.3333	Low
39	0.3333	Low
40	0.375	Low
41	0.4348	Moderate
42	0.375	Low
43	0.3889	Low
44	0.3846	Low
45	0.4	Low
46	0.4483	Moderate
47	0.4407	Moderate
48	0.3333	Low
49	0.3333	Low
50	0.4222	Moderate
51	0.3333	Low
52	0.3333	Low
53	0.4222	Moderate
54	0.4565	Moderate
55	0.4596	Moderate
56	0.3636	Low
57	0.4	Low
58	0.4091	Moderate
59	0.3846	Low
60	0.3939	Low
61	0.4167	Moderate
62	0.4	Low
63	0.4333	Moderate

<b>student</b>	<b>ALA</b>	<b>Classification</b>
64	0.4	Low
65	0.4642	Moderate
66	0.3333	Low
67	0.4	Low
68	0.4464	Moderate
69	0.4	Low
70	0.4	Low
71	0.3333	Low
72	0.3333	Low
73	0.4138	Moderate
74	0.4	Low
75	0.45	Moderate
76	0.4375	Moderate
77	0.4681	Moderate
78	0.45	Moderate
79	0.4118	Moderate
80	0.4486	Moderate
81	0.45	Moderate
82	0.3846	Low
83	0.4355	Moderate
84	0.4258	Moderate
85	0.4435	Moderate
86	0.3333	Low
87	0.4	Low
88	0.381	Low
89	0.4375	Moderate
90	0.4328	Moderate
91	0.375	Low
92	0.4	Low
93	0.4576	Moderate
94	0.4	Low
95	0.4103	Moderate
96	0.4	Low
97	0.4167	Moderate
98	0.4444	Moderate
99	0.4255	Moderate
100	0.4167	Moderate
101	0.4219	Moderate
102	0.375	Low
103	0.4	Low
104	0.3333	Low
105	0.3333	Low

<b>student</b>	<b>ALA</b>	<b>Classification</b>
106	0.449	Moderate
107	0.4149	Moderate
108	0.4465	Moderate
109	0.4615	Moderate
110	0.4	Low
111	0.4842	Moderate
112	0.4286	Moderate
113	0.4626	Moderate
114	0.4643	Moderate
115	0.4706	Moderate
116	0.466	Moderate
117	0.4286	Moderate
118	0.4588	Moderate
119	0.4689	Moderate
120	0.4324	Moderate
121	0.3333	Low
122	0.4	Low
123	0.47	Moderate
124	0.4	Low
125	0.4528	Moderate
126	0.3333	Low
127	0.4697	Moderate
128	0.4706	Moderate
129	0.4519	Moderate
130	0.3333	Low
131	0.4318	Moderate
132	0.4833	Moderate
133	0.4673	Moderate
134	0.457	Moderate
135	0.3333	Low
136	0.4882	Moderate
137	0.4359	Moderate
138	0.4746	Moderate
139	0.4464	Moderate
140	0.4286	Moderate
141	0.4	Low
142	0.4592	Moderate
143	0.4414	Moderate
144	0.4597	Moderate
145	0.4688	Moderate
146	0.4364	Moderate
147	0.4669	Moderate

<b>student</b>	<b>ALA</b>	<b>Classification</b>
148	0.4757	Moderate
149	0.4694	Moderate
150	0.4706	Moderate
151	0.4737	Moderate
152	0.4862	Moderate
153	0.4664	Moderate
154	0.438	Moderate
155	0.4815	Moderate
156	0.48	Moderate
157	0.4545	Moderate
158	0.4046	Moderate
159	0.4353	Moderate
160	0.4734	Moderate
161	0.3333	Low
162	0.4769	Moderate
163	0.475	Moderate
164	0.3333	Low
165	0.4873	Moderate
166	0.4694	Moderate
167	0.4781	Moderate
168	0.4828	Moderate
169	0.4854	Moderate
170	0.4952	Moderate
171	0.4457	Moderate
172	0.4704	Moderate
173	0.475	Moderate
174	0.4706	Moderate
175	0.4591	Moderate
176	0.4913	Moderate
177	0.496	Moderate
178	0.4808	Moderate
179	0.4901	Moderate
180	0.4884	Moderate
181	0.4634	Moderate
182	0.4286	Moderate
183	0.4483	Moderate
184	0.4	Low
185	0.4948	Moderate
186	0.4873	Moderate
187	0.4811	Moderate
188	0.4913	Moderate
189	0.3333	Low



student	ALA	Classification
190	0.4851	Moderate
191	0.4792	Moderate
192	0.4286	Moderate
193	0.4879	Moderate
194	0.3333	Low
195	0.3333	Low
196	0.4828	Moderate
197	0.4667	Moderate
198	0.4888	Moderate
199	0.4921	Moderate
200	0.4774	Moderate
201	0.4762	Moderate
202	0.4904	Moderate
203	0.4875	Moderate
204	0.4528	Moderate
205	0.4936	Moderate
206	0.439	Moderate
207	0.492	Moderate
208	0.4444	Moderate
209	0.4	Low
210	0.4965	Moderate
211	0.4	Low
212	0.3333	Low
213	0.4941	Moderate
214	0.4737	Moderate
215	0.4981	Moderate
216	0.495	Moderate
217	0.4	Low
218	0.4904	Moderate
219	0.4979	Moderate

*Appendix 28. Classification of results predicted for IRA*

Student	IRA	Class
1	0.81761	High
2	0.550847	moderate
3	0.321833	Low
4	0.312625	Low
5	0.557971	moderate
6	0.4051	Low
7	0.333288	Low
8	0.527869	moderate

Student	IRA	Class
9	0.779874	High
10	0.521994	moderate
11	0.766513	High
12	0.445917	Low
13	0.402359	Low
14	0.686567	moderate
15	0.608696	moderate
16	0.661871	moderate
17	0.732558	High
18	0.704981	High
19	0.594444	moderate
20	0.375281	Low
21	0.671429	moderate
22	0.612195	moderate
23	0.305556	Low
24	0.834532	High
25	0.725363	High
26	0.874214	High
27	0.761224	High
28	0.679245	moderate
29	0.336498	Low
30	0.393233	Low
31	0.743764	High
32	0.811321	High
33	0.615686	moderate
34	0.688462	moderate
35	0.541401	moderate
36	0.786164	High
37	0.807843	High
38	0.693467	moderate
39	0.665306	moderate
40	0.821918	High
41	0.791111	High
42	0.794297	High
43	0.868597	High
44	0.687646	moderate
45	0.777941	High
46	0.604502	moderate
47	0.832402	High
48	0.827338	High
49	0.765625	High

Student	IRA	Class
50	0.631111	moderate
51	0.758701	High
52	0.587755	moderate
53	0.5125	moderate
54	0.793434	High
55	0.784746	High
56	0.565674	moderate
57	0.741784	High
58	0.612708	moderate
59	0.783231	High
60	0.808081	High
61	0.776471	High
62	0.628302	moderate
63	0.746562	High
64	0.701961	High
65	0.682979	moderate
66	0.685499	moderate
67	0.642553	moderate
68	0.648485	moderate
69	0.757856	High
70	0.825532	High
71	0.595745	moderate
72	0.637143	moderate
73	0.827338	High
74	0.502924	moderate
75	0.837375	High
76	0.665138	moderate
77	0.776978	High
78	0.636616	moderate
79	0.712251	High
80	0.346296	Low
81	0.233756	Low
82	0.837416	High
83	0.681579	moderate
84	0.807018	High
85	0.601719	moderate
86	0.593939	moderate
87	0.710909	High
88	0.613596	moderate
89	0.73	High
90	0.552174	moderate

Student	IRA	Class
91	0.756667	High
92	0.515517	moderate
93	0.771689	High
94	0.649326	moderate
95	0.625352	moderate
96	0.745575	High
97	0.765823	High
98	0.636232	moderate
99	0.581481	moderate
100	0.745572	High
101	0.680645	moderate
102	0.779412	High
103	0.672087	moderate
104	0.538951	moderate
105	0.858537	High
106	0.658065	moderate
107	0.728571	High
108	0.574388	moderate
109	0.750542	High
110	0.732323	High
111	0.574359	moderate
112	0.661224	moderate
113	0.78744	High
114	0.503125	moderate
115	0.562162	moderate
116	0.621716	moderate

*Appendix 29. Classification of results predicted for the IPS*

Student	IPS	Classification
1	0.771689	High
2	0.661224	Moderate
3	0.587755	Moderate
4	0.621716	Moderate
5	0.78744	High
6	0.503125	Moderate
7	0.649326	Moderate
8	0.574359	Moderate
9	0.562162	Moderate
10	0.637143	Moderate
11	0.776978	High

Student	IPS	Classification
12	0.538951	Moderate
13	0.765823	High
14	0.672087	Moderate
15	0.515517	Moderate
16	0.766513	High
17	0.612708	Moderate
18	0.625352	Moderate
19	0.636232	Moderate
20	0.527869	Moderate
21	0.594444	Moderate
22	0.827338	High
23	0.834532	High
24	0.737288	High
25	0.658065	Moderate
26	0.712625	High
27	0.601719	Moderate
28	0.615686	Moderate
29	0.665306	Moderate
30	0.921833	High
31	0.736498	High
32	0.701961	High
33	0.693467	Moderate
34	0.728571	High
35	0.680645	Moderate
36	0.5125	Moderate
37	0.581481	Moderate
38	0.574388	Moderate
39	0.745572	High
40	0.761224	High
41	0.7051	High
42	0.631111	Moderate
43	0.732323	High
44	0.807843	High
45	0.837416	High
46	0.868597	High
47	0.685499	Moderate
48	0.733756	High
49	0.858537	High
50	0.825532	High
51	0.705556	High
52	0.777941	High
53	0.741784	High

Student	IPS	Classification
54	0.776471	High
55	0.732558	High
56	0.841176	High
57	0.791111	High
58	0.557971	Moderate
59	0.686567	Moderate
60	0.642553	Moderate
61	0.682979	Moderate
62	0.552174	Moderate
63	0.688462	Moderate
64	0.716667	High
65	0.704981	High
66	0.758701	High
67	0.756667	High
68	0.793434	High
69	0.608696	Moderate
70	0.784746	High
71	0.541401	Moderate
72	0.613596	Moderate
73	0.521994	Moderate
74	0.595745	Moderate
75	0.679245	Moderate
76	0.750542	High
77	0.710909	High
78	0.807018	High
79	0.743764	High
80	0.550847	Moderate
81	0.725363	High
82	0.825	High
83	0.712251	High
84	0.827338	High
85	0.604502	Moderate
86	0.746562	High
87	0.765625	High
88	0.687646	Moderate
89	0.612195	Moderate
90	0.628302	Moderate
91	0.779412	High
92	0.757856	High
93	0.702359	High
94	0.779874	High
95	0.745917	High

Student	IPS	Classification
96	0.746296	High
97	0.793233	High
98	0.775281	High
99	0.808081	High
100	0.636616	Moderate
101	0.811321	High
102	0.565674	Moderate
103	0.73	High
104	0.648485	Moderate
105	0.745575	High
106	0.81761	High
107	0.661871	Moderate
108	0.665138	Moderate
109	0.783231	High
110	0.874214	High
111	0.681579	Moderate
112	0.794297	High
113	0.837375	High
114	0.786164	High
115	0.593939	Moderate
116	0.832402	High
117	0.821918	High
118	0.671429	Moderate
119	0.502924	Moderate
120	0.7	High

*Appendix 30. Classification of results predicted for the WMC.*

student	WMC	classification
1	0.999834	High
2	0.999636	High
3	0.999417	High
4	0.999556	High
5	1	High
6	0.995819	High
7	0.972569	High
8	0.98337	High
9	0.992428	High
10	0.994605	High
11	0.975739	High
12	0.995772	High
13	0.990989	High

<b>student</b>	<b>WMC</b>	<b>classification</b>
14	0.98027	High
15	0.997559	High
16	0.987828	High
17	0.998021	High
18	0.994348	High
19	0.991612	High
20	0.938653	High
21	0.977602	High
22	0.873982	High
23	0.987761	High
24	0.898652	High
25	0.965757	High
26	0.988174	High
27	0.994142	High
28	0.951764	High
29	0.938395	High
30	0.979629	High
31	0.939671	High
32	0.926471	High
33	0.986476	High
34	0.907701	High
35	0.984781	High
36	0.984202	High
37	0.963046	High
38	0.992532	High
39	0.996428	High
40	0.978481	High
41	0.72728	High
42	0.948832	High
43	0.728931	High
44	0.761716	High
45	0.840278	High
46	0.974939	High
47	0.966107	High
48	0.962631	High
49	0.697332	Moderate
50	0.553996	Moderate
51	0.910667	High
52	0.993671	High
53	0.979564	High
54	0.874021	High
55	0.937226	High



<b>student</b>	<b>WMC</b>	<b>classification</b>
56	0.544267	Moderate
57	0.987191	High
58	0.85034	High
59	0.94929	High
60	0.982683	High
61	0.943653	High
62	0.917129	High
63	0.741787	High
64	0.924054	High
65	0.922085	High
66	0.960753	High
67	0.985057	High
68	0.648625	Moderate
69	0.9098	High
70	0.989838	High
71	0.956007	High
72	0.914899	High
73	0.833763	High
74	0.92446	High
75	0.729167	High
76	0.953007	High
77	0.692989	Moderate
78	0.894431	High
79	0.930328	High
80	0.619552	Moderate
81	0.880923	High
82	0.866759	High
83	0.839348	High
84	0.911237	High
85	0.696769	Moderate
86	0.854355	High
87	0.957199	High
88	0.958237	High
89	0.670906	Moderate
90	0.732388	High
91	0.734605	High
92	0.414634	Low
93	0.932	High
94	0.568627	Moderate
95	0.396899	Low
96	0.964765	High
97	0.430939	Low

<b>student</b>	<b>WMC</b>	<b>classification</b>
98	0.899922	High
99	0.826393	High
100	0.505004	Moderate
101	0.805177	High
102	0.13624	Low
103	0.703393	High
104	0.761463	High
105	0.822047	High
106	0.732406	High
107	0.760683	High
108	0.747276	High
109	0.38174	Low
110	0.725806	High
111	0.263692	Low
112	0.269647	Low
113	0.743226	High
114	0.683521	Moderate
115	0.476786	Low
116	0.476047	Low
117	0.766197	High
118	0.908127	High
119	0.681602	Moderate
120	0.718398	High
121	0.780054	High
122	0.218834	Low
123	0.643253	Moderate
124	0.45125	Low
125	0.804498	High
126	0.701129	High
127	0.236358	Low
128	0.657982	Moderate
129	0.830384	High
130	0.865854	High
131	0.739486	High
132	0.629447	Moderate
133	0.786885	High
134	0.845672	High
135	0.088235	Low
136	0.966459	High
137	0.397315	Low
138	0.322581	Low
139	0.586185	Moderate

student	WMC	classification
140	0.557084	Moderate
141	0.376344	Low
142	0.084888	Low
143	0.243017	Low
144	0.288106	Low
145	0.438247	Low
146	0.391371	Low
147	0.466014	Low
148	0.532795	Moderate
149	0.387283	Low
150	0.531053	Moderate
151	0.17757	Low
152	0.557339	Moderate
153	0.141955	Low

*Appendix 31. Sample Training Dataset for the ANN and the cluster models.*

The columns ABC, D, AAC, A, and V\_S indicate the number of times a student viewed pages with relevant content. ABC\_T, D\_T, AAC\_T, and A\_T, indicate the time in minutes a student spent on pages with relevant contents. Ascore, Attempts, and AVG stand for marks scored on online tests, the number of attempts, and the average content covered respectively.

Userid	ABC	D	C	AAC	A	V_S	ABC_T	D_T	C_T	AAC_T	A_T	Ascore	Attempts	AVG	LS
280774	86	60	20	80	0	50	80	297	560	156	0	65	1	59	3
280775	71	100	80	100	100	100	11	48	144	15	36	66	1	75	5
280776	100	100	100	80	100	100	16	51	150	16	0	81	1	61	8
280777	100	60	20	100	0	50	75	99	440	19	0	58	1	91	9
280778	57	80	21	40	0	100	8	73	120	6	0	61	1	21	7
280779	100	100	60	100	100	100	10	26	74	12	12	60	2	73	12
280780	71	100	60	60	100	100	15	48	239	25	15	62	1	85	13
280781	71	80	80	80	100	100	28	91	243	20	33	49	2	49	15
280782	71	80	80	100	100	100	14	36	96	17	18	66	1	47	17
280783	71	40	40	40	100	100	17	35	107	6	22	69	1	27	18
280784	100	80	80	80	0	100	25	87	164	27	0	78	1	15	19
280785	86	100	80	80	100	100	5	15	56	5	12	69	1	47	17
280786	71	60	60	100	100	50	13	75	216	50	33	72	1	28	19
280787	71	60	0	40	100	100	36	149	300	25	49	70	1	11	21
280788	71	40	21	60	100	50	105	92	302	12	55	74	1	42	21
280789	86	100	21	60	100	50	13	17	48	2	6	73	1	43	22

*Appendix 32. Neural Network Program Code Segment.*

## **This is R language program code for neural network analysis to predict learning styles and cognitive traits**

### **Setting a working directory with the dataset**

```
setwd("C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-  
studio/NeuralNetwork/")
```

```
mydata <- read.csv("LS_CT_05_06_2019.csv")
```

```
attach(mydata)
```

```
str(mydata)
```

```
data=mydata
```

### **Scaling dataset**

```
max_data <- apply(data[,1:14], 2, max)
```

```
min_data <- apply(data[,1:14], 2, min)
```

```
data_scaled <- scale(data[,1:14],center = min_data, scale = max_data - min_data)
```

```
LS = as.numeric(data$LS)-1
```

```
data_scaled = cbind(LS,data_scaled)
```

### **Splitting dataset into training (70%) and testing set (30%)**

```
index = sample(1:nrow(data),round(0.70*nrow(data)))
```

```
train_data <- as.data.frame(data_scaled[index,])
```

```
test_data <- as.data.frame(data_scaled[-index,])
```

### **Adding neural network library**

```
library("neuralnet")
```

```
library(NeuralNetTools)
```

```
library(RSNNS)
```

### **Training the network**

```
n = names(train_data)
```

```
f <- as.formula(paste("LS ~", paste(n[!n %in% "LS"], collapse = " + ")))
```

```
deep_net = neuralnet(f,data=train_data,hidden=c(7,4,3,2,1),linear.output=F, threshold =
0.001,learningrate = NULL,learningrate.factor = list(minus = 0.5,plus = 1.2), stepmax = 1e+05,
rep = 10)
```

### **Plotting the network**

```
plot(deep_net,arrow.length = 0.15)
```

```
x <- train_data[,
c('ABC','D','C','AAC','A','V_S','ABC_T','D_T','C_T','AAC_T','A_T','Ascore','Attempts','AVG')]
```

```
y <- train_data[, 'LS']
```

```
mod <- mlp(x, y, size = 11)
```

```
# garson
```

```
garson(mod, LS)
```

```
#garson(deep_net)
```

### **Displaying confusion matrix**

```
predicted_data <- compute(deep_net,test_data[,1:14])
```

```
print(head(predicted_data$net.result))
```

```
predicted_data$net.result <- sapply(predicted_data$net.result,round,digits=0)
```

```
results <- data.frame(actual = test_data$LS, prediction = predicted_data$net.result)
```

```
library(caret)
```

```
results$prediction <- as.factor(sample(c(0, 1), 93, replace=T))
```

```
results$actual <- as.factor(sample(c(0, 1), 93, replace=T))
```

```
result <- caret::confusionMatrix(results$prediction, results$actual, mode="prec_recall")
```

```
result
```

### **plotting ROC CURVE**

```
actual<-results$actual
```

```
prediction <- rev(seq_along(actual))
```

```

prediction[9:10] <- mean(prediction[9:10])
library(pROC)
roc_obj <- roc(actual,prediction)
auc(roc_obj)
roc_df <- data.frame(
  TPR=rev(roc_obj$sensitivities),
  FPR=rev(1 - roc_obj$specificities),
  labels=roc_obj$response,
  scores=roc_obj$predictor)
rectangle <- function(x, y, width, height, density=12, angle=-45, ...)
  polygon(c(x,x,x+width,x+width), c(y,y+height,y+height,y),
    density=density, angle=angle, ...)
roc_df <- transform(roc_df,
  dFPR = c(diff(FPR), 0),
  dTPR = c(diff(TPR), 0))
plot(0:10/10, 0:10/10, type='n', xlab="FPR", ylab="TPR")
abline(h=0:10/10, col="lightblue")
abline(v=0:10/10, col="lightblue")
with(roc_df, {
  mapply(rectangle, x=FPR, y=0,
    width=dFPR, height=TPR, col="green", lwd=2)
  mapply(rectangle, x=FPR, y=TPR,
    width=dFPR, height=dTPR, col="blue", lwd=2)
  lines(FPR, TPR, type='b', lwd=3, col="red")
})

```

```
#Adding up the area
simple_auc <- function(TPR, FPR){
  # inputs already sorted, best scores first
  dFPR <- c(diff(FPR), 0)
  dTPR <- c(diff(TPR), 0)
  sum(TPR * dFPR) + sum(dTPR * dFPR)/2 }
with(roc_df, simple_auc(TPR, FPR))
```

*Appendix 33. Sample Prediction Results for the ANN model.*

STUDENT	Actual	Prediction	
1	0	0	
2	0	0	
3	1	0	
4	1	0	
5	1	1	
6	0	0	
7	0	0	
10	0	0	
12	0	0	
13	1	1	
14	1	0	
15	0	0	
16	1	0	
17	0	0	
18	0	0	
19	0	1	
20	1	1	
21	1	0	
22	1	1	

*Appendix 34. Cluster Analysis Program Code Segment.*

**STEP 1: LOADING LIBRARIES**

```
library("factoextra")
```

## **STEP 2: LOADING DATASET**

```
df <- read.csv("C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-  
studio/cluster/exercises/LS_EXPERIMENT_14_03_2020.csv",header=T)
```

## **PREPROCESSING**

### **# a.removing missing values**

```
df <- na.omit(df)
```

```
my_data=df
```

```
str(my_data)
```

### **#b. scaling variables**

```
#my_data <- scale(my_data[c(2:16)])
```

```
my_data<-scale(my_data[,2:7])
```

```
#display first 10 records
```

```
head(my_data)
```

### **# c. Assessing the clusterability using the Hopkins statistic method**

```
library("factoextra")
```

```
res <- get_clust_tendency(my_data, 300, graph = TRUE)
```

```
res$hopkins_stat
```

```
print(res$plot)
```

```
res <- get_clust_tendency(my_data, n = nrow(df)-1, graph = FALSE)
```

```
res$hopkins_stat
```

### **# Compute Hopkins statistics for the LS dataset**

```
res <- get_clust_tendency(my_data, n = nrow(my_data)-1, graph = FALSE)
```

```
res$hopkins_stat
```

### **# visualizing assessment of cluster tendency (VAT) approach**



```
print(res$plot)
```

```
head(my_data, n = 3)
```

#### **#d. Clarifying distance measures**

```
res.dist <- get_dist(my_data, stand = TRUE, method = "pearson")
```

```
fviz_dist(res.dist,
```

```
  gradient = list(low = "#00AFBB", mid = "white", high = "#FC4E07"))
```

### **#STEP 3. DETERMINE THE OPTIMAL NUMBER OF CLUSTERS**

```
#Determine the optimal number of clusters: use factoextra::fviz_nbclust()
```

```
library(NbClust)
```

```
library("factoextra")
```

#### **# method 1 Elbow method**

```
fviz_nbclust(my_data, kmeans, method = "wss") +
```

```
  geom_vline(xintercept = 4, linetype = 2)+
```

```
  labs(subtitle = "Elbow method")
```

#### **# method 2 Silhouette method**

```
fviz_nbclust(my_data, kmeans, method = "silhouette")+
```

```
  labs(subtitle = "Silhouette method")
```

#### **#method 3 Gap statistic method**

```
set.seed(123)
```

```
fviz_nbclust(my_data, kmeans, nstart = 25, method = "gap_stat", nboot = 50)+
```

```
  labs(subtitle = "Gap statistic method")
```

#### **#method 4 NbClust**

```
set.seed(123)
```

```
res.nbclust <- NbClust(my_data, distance = "euclidean",
```

```
  min.nc = 2, max.nc = 15,
```

```

        method = "complete", index = "all")

#Visualize using factoextra:

factoextra::fviz_nbclust(res.nbclust) + theme_minimal()

#STEP 4: CLUSTERING

# method 1: Compute and visualize k-means clustering

set.seed(123)
km.res<-kmeans(my_data, 6, iter.max = 100, nstart = 25)
print(km.res)
# process distribution of clusters by learning styles
k.means.fit<-km.res
c_ls_data<-table(df$LS,k.means.fit$cluster)
c_ls_data
#write the results on a CSV file

write.csv(c_ls_data,"C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-
studio/cluster/training_output/kmeans_distribution_LS_EXPERIMENT_14_03_2020_LS_.csv")

#list students by clusters

c_data=table(df[,1],km.res$cluster)

c_data

##write the results on a CSV file

write.csv(c_data,"C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-
studio/cluster/training_output/kmeans_LS_EXPERIMENT_GROUPS.csv")

# Visualize cluster

library("factoextra")

fviz_cluster(km.res, data = my_data, frame.type = "ellipse")+ theme_minimal()

#method 2: Compute and visualize PAM clustering (Partitioning Around Medoids).

library("cluster")
pam.res <- pam(my_data, 6)
print(pam.res)

# Visualize

```

```
fviz_cluster(pam.res)
```

### **# process distribution of clusters by learning styles**

```
pam.res.fit<-pam.res  
c_p_data<-table(df$LS,pam.res.fit$cluster)  
c_p_data
```

### **#write CSV distribution file**

```
write.csv(c_p_data,"C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-  
studio/cluster/training_output/distribution_LS_EXPERIMENT_14_03_2020_PAM.csv")
```

### **# display students list by Partitioning Around Medoids**

```
p_data=table(df[,1],pam.res$cluster)  
p_data  
write.csv(p_data,"C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-  
studio/cluster/training_output/PAM_LS_EXPERIMENT_GROUPS.csv")
```

### **#method 3: Compute and visualize Hierarchical clustering**

```
d <- dist(my_data, method = "euclidean")  
res.hc <- hclust(d, method = "ward.D2" )  
grp <- cutree(res.hc, k = 6)
```

### **# process distribution of clusters by learning styles**

```
g_data<-table(df$LS,grp)  
g_data  
#write CSV distribution file  
write.csv(g_data,"C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-  
studio/cluster/training_output/HIERACHY_LS_EXPERIMENT_14_03_2020.csv")  
t_data=table(df[,1],grp)  
t_data
```

### **#write CSV distribution file**

```
write.csv(t_data,"C:/Users/lwandeomondi/Desktop/research/CHAPTERS/ML COURSE/r-  
studio/cluster/training_output/HIERACHY_LS_EXPERIMENT_GROUPS.csv")
```

### **# Visualize**

```
plot(res.hc, cex = 1)  
rect.hclust(res.hc, k = 6, border = 2:20) # add rectangle
```

```

library("factoextra")

res <- hcut(my_data, k = 6, stand = TRUE)

fviz_dend(res, rect = TRUE, cex = 0.5,
           k_colors = c("#00AFBB", "#DC143C",
                        "#7CFC00", "#FA8072", "#228B22", "#0000CD", "#FF8C00", "#8B0000", "#6495ED", "#483D8B"
                        ))

```

## **STEP 5: CLUSTER EVALUATION**

METHOD 1. silhouette analysis

**# Silhouette coefficient of observations**

**# display summary of silhouette analysis**

```
si.sum <- summary(sil)
```

```
si.sum
```

**Silhouette plot for k-means**

```
plot(sil, main = "Silhouette plot - K-means")
```

**Silhouette plot for k-means clustering**

```
fviz_silhouette(km.res)
```

**Silhouette plot for PAM clustering**

```
fviz_silhouette(pam.res)
```

**Silhouette plot for hierarchical clustering**

```
fviz_silhouette(res.hc)
```

## **#STEP 6: DISPLAY CLUSTER STATISTICS**

**#Cluster statistics for k-means clustering**

```
library(fpc)
```

**# Compute pairwise-distance matrices**

```
dd <- dist(my_data, method = "euclidean")
```

```

# Statistics for k-means clustering
km_stats <- cluster.stats(dd, km.res$cluster)

# (k-means) within clusters sum of squares
km_stats$within.cluster.ss

# (k-means) cluster average silhouette widths
km_stats$clus.avg.silwidths

# Display all statistics
km_stats

#Cluster statistics for PAM clustering
pam_stats <- cluster.stats(dd, pam.res$cluster)

# (pam) within clusters sum of squares
pam_stats$within.cluster.ss

# (pam) cluster average silhouette widths
pam_stats$clus.avg.silwidths

#Cluster statistics for hierarchical clustering
hc_stats <- cluster.stats(dd, res$cluster)

# (HCLUST) within clusters sum of squares
hc_stats$within.cluster.ss

# (HCLUST) cluster average silhouette widths
hc_stats$clus.avg.silwidths

```

Appendix 35. Sample Cluster model results for 20 Students.

Student	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	1	0	0
2	0	0	0	0	0	0	1	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0	1	0	0
6	1	0	0	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	1	0	0
10	0	0	0	0	0	0	0	0	0	1
11	0	0	1	0	0	0	0	0	0	0
12	0	0	1	0	0	0	0	0	0	0
13	0	0	0	1	0	0	0	0	0	0
14	0	0	1	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	1	0	0
16	0	1	0	0	0	0	0	0	0	0
17	0	0	0	1	0	0	0	0	0	0
18	0	1	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	1	0
20	0	0	0	0	0	1	0	0	0	0

Appendix 36. Pattern extraction queries

extracting patterns related to active LS

```
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user, cl_log, HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress, HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
and name like '%activit%'
group by user_id
```

extracting patterns related to reflective LS

```
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user, cl_log, HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress, HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
and name like '%summary%'
or name like '%self-assessment%'
```

```

or name like '%revision%'
group by user_id
extracting patterns related to sensing LS
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user, cl_log, HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress, HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
and name like '%1.4%'
or name like '%2.4%'
or name like '%3.4%'
and name not like '%summary%'
and name not like '%self-assessment%'
and name not like '%revision%'
group by user_id
extracting patterns related to intuitive LS
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user, cl_log, HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress, HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
and name like '%1.3%'
or name like '%2.3%'
or name like '%3.3%'
group by user_id
extracting patterns related to global LS
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user, cl_log, HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress, HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
and name like '%unit%'
or name like '%section%' or name like '%module%' group by user_id

extracting patterns related to visual LS
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user, cl_log, HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress, HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
and name like '%figure%'
or name like '%picture%'
or name like '%illustration%'
group by user_id

```

```

extracting patterns related to AIA
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user,cl_log,HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress,HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
group by user_id, date having count(date)>1
extracting patterns related to WMC
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user,cl_log,HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress,HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
group by user_id, date having count(date)=1
extracting patterns related to IRA
select user_id, count(name), sec_to_time(sum(total_time)) from
cl_user,cl_log,HCH100_lp_module,
HCH100_lp_rel_learnPath_module,
HCH100_lp_user_module_progress,HCH100_lp_asset
where cl_user.user_id=HCH100_lp_user_module_progress.user_id
and HCH100_lp_module.module_id=HCH100_lp_rel_learnPath_module.module_id
and HCH100_lp_asset.module_id=HCH100_lp_rel_learnPath_module.module_id
and cl_user.user_id=cl_log.user_id
and name like "%example%" or
and name like "%case stud%"
group by user_id
extracting patterns related to IPS
the select result,time,attempts,
from cl_user_id,c_HCH100_qwz_tracking where
cl_user.user_id=c_HCH100_qwz_tracking.user_id
group by user_id

```