ATHABASCA UNIVERSITY

TOWARDS A PERSONALIZED STUDY GUIDE BASED ON BEHAVIOUR ANALYTICS

BY

THEODORE KRAHN

A THESIS/DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN INFORMATION SYSTEMS

DR. MAIGA CHANG

ATHABASCA, ALBERTA

NOVEMBER 2024

Athabasca University
Faculty of Graduate Studies

Approval of Thesis

The undersigned certify that they have read the thesis entitled

PERSONALIZED STUDY GUIDE GENERATION BASED ON BEHAVIOUR ANALYTICS

Submitted by

Theodore Krahn

In partial fulfillment of the requirements for the degree of

Master of Science in Information Systems

The thesis examination committee certifies that the thesis
and the oral examination is approved

Supervisor:
Dr. Maiga Chang
Athabasca University

Committee Members:
Dr. Farook Al-Shamali
Athabasca University

Dr. Ahmed Tlili
Beijing Normal University

Dr. Rita Kuo
Utah Valley University

External Examiner:
Dr. Tho Pham Duc
Vietnam National University

December 20, 2024

1 University Drive, Athabasca, AB, T9S 3A3 Canada
Toll-free (CAN/U.S.) 1.800.788.9041 ex. 6821
fgs@athabascau.ca | fgs.athabascau.ca | athabascau.ca

Abstract

Personalized learning may help students achieve better learning outcomes. To provide personalization, the student preferences must be known. Learning styles determine preferences and detecting learning style automatically rather than using a questionnaire is the topic of much research. This research proposes another technique to non-invasively elicit learning styles. The proposed technique will construct a network graph from the learning objects in a course, where each node in the graph represents a learning object. The nodes will be positioned using the similarity between learning object contents as the distance between nodes. Then, student behaviour in the form of log traces will be added as links to the graph, producing a visualization of student behaviour, named Learning Object Graph (LOG). Each student will have their behaviour compressed into a centroid point for clustering. The resulting student groups will be classified according to their learning style so study guide personalization can be done.

Keywords: learning style, personalized learning, log file clustering, learning analytics, sequential behaviour

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter 1. Introduction

1.1 Motivation and Goal

In the field of personalized and adaptive learning it is idealized that each student should have a learning experience based on their own unique preferences or abilities. While it may be difficult to ascertain an individual's unique preferences or abilities, there may be groups among individuals whose preferences are easier to determine. Tailoring learning material to groups of individuals will be easier than trying to customize content for each unique person, while still providing the benefit of a personalized learning experience. In turn, a personalized learning experience can help the student to work through the course material and better understand the course content. If the student has a personalized learning experience, it follows that their learning outcome will be superior to a generic learning experience (Pane et al., 2017).

The problem is to understand the differences and similarities between groups of students so that the learning material can be effectively customized for each group. One possible method to group students is by their learning styles, which is the most commonly modeled characteristic in other research (Martin et al., 2020; Afini Normadhi et al., 2019). There are many different learning style models, but the Felder-Silverman Learning Style Model (FSLSM) is the most used (Afini Normadhi et al., 2019). According to the FSLSM:

> "Students learn in many ways— by seeing and hearing; reflecting and acting; reasoning logically and intuitively; memorizing and visualizing and drawing analogies and building mathematical models; steadily and in fits and starts." (Felder & Silverman, 1988)

There is considerable debate about whether learning styles from any model are actually real and useful. While some postulate that learning styles are not worth considering, there is still much research being done utilizing learning styles, as this research does. It is hoped that the outcome of this research will bolster the argument for considering learning styles in personalized learning.

However, most learning style models determine a student's style with questionnaires, such as the FSLSM which uses 44 questions to group students into 16 possible learning styles. These tests are time consuming to administer and must be done for each student. A mechanism that could automatically determine the student's learning style from their behaviour without the use of a questionnaire would be preferable.

If students' learning styles can be determined from their behaviour when interacting with the learning system, then the students can be grouped. The learning content can be organized for each student based on their learning style group and provide each student with a personalized learning experience.

1.2 Research Purpose

To personalize learning content, the student's learning preferences or learning style must be determined. Learning style questionnaires are the traditional approach to determining learning style. However, lengthy questionnaires are time consuming for the teacher to administer and for the student to take. Furthermore, students may get bored when filling out the questionnaire and not consider the importance of the task, leading them to give arbitrary answers and invalidating the results. Additionally, some learning style theories suggest that learning style is dynamic, changing in response to the learning task at hand (Truong, 2016; Wu et al., 2017). Under these circumstances, a static questionnaire, given once, fails to capture the dynamic nature of the student's learning process.

Eliciting student learning styles from their behaviour is an alternative to taking a static one-time questionnaire. In an e-learning environment, the student behaviour is captured in the logs of the learning management system (LMS), which can be extracted and analyzed. The problem is in mapping the behaviour to a learning style and there have been other attempts at resolving this mapping (Aissaoui et al., 2019a; Dung & Florea, 2012).

One technique that has been used is to cluster the log traces, which can be done in a few different ways. First, the various learning objects in a course are mapped to different dimensions of the learning style theory. Then, a clustering algorithm can be used to group students according to which learning objects they accessed. This is the approach taken by Aissaoui et al. (2019a), which mapped the learning objects to the Felder-Silverman learning style model (FSLSM), then constructed feature vectors from student sessions, and used k-modes clustering to group the students.

The same authors did similar research where they compared a fuzzy clustering algorithm, where students could be in more than one group, against k-means, concluding that fuzzy clustering is the superior of the two (Aissaoui et al., 2019b). Fuzzy clustering makes sense from the standpoint that a student might be borderline between different learning styles. On the other hand, if a student is labeled as having multiple learning styles, then which classification should be considered when providing personalization?

Dung & Florea (2012) take a similar approach to learning style detection, where each learning object in the course is classified as belonging to a particular dimension of the FSLSM. These authors also consider the time spent with a learning object when grouping students. Li & Abdul Rahman (2018) also map learning objects to the FSLSM and compared a tree-augmented naive Bayesian network to a regular Bayesian network, concluding that the tree-augmented version was able to classify learning styles more accurately.

The primary benefit of the above-mentioned approaches is that they allow for a student to be classified differently at different times, which considers the dynamic nature of learning styles. When there are different types of learning objects with the same content, a video and transcript of a lecture for instance, it makes sense to consider the student a visual learner when they watch the video instead of reading the transcript. However, many courses do not contain different learning objects with the same content and can also contain learning objects that are of a similar kind, perhaps all reading material, which makes the mapping less meaningful.

The use of clustering algorithms with log data often takes a direct approach and uses the log traces either directly or in a slightly modified form. Both Anandhi & Ahmed (2019) and Valsamidis et al. (2012) clean the log data to remove outliers and noise, then user sessions are determined, and finally clustering is done. Anandhi & Ahmed (2019) used the log traces directly, while Valsamidis et al. (2012) also created extra features such as enrichment, disappointment, and interest using metrics calculated from the log data.

Clustering with sessions makes sense when different users can be expected to have similar sessions. But, in distance learning, some students are full-time, while others are part-time, some study a couple hours a day, while others study all day. These students will have sessions that are not always comparable. Also, other research that examined clustering log traces does not provide any visualization of the process. The clustering algorithm will run and produce classifications, but the behaviour patterns are not visualized.

The purpose of this research is to provide a novel log clustering-based approach to learning style detection that ignores sessions and provides a visualization of both the student behaviour patterns and the clustering results in the form of the Behaviour Analytics plugin. Considering student data in aggregate rather than by session is appropriate in a distance learning environment and the ability to visualize behaviour patterns is something most other research have not done.

The purpose of this research is also to provide an enhanced learning experience through the Personalized Study Guide plugin, which will use the learning style groups from the Behaviour Analytics plugin. Once the student learning styles are determined from student behaviour, the study guide can be personalized for each student's style, thereby providing an enhanced learning experience.

Plugins for the Moodle Learning Management System (LMS) will be developed to provide the learning style detection mechanism and to provide the personalized learning experience. Moodle has been selected because it is freely licensed under the GNU GPL (Free Software Foundation, 2007) and is the LMS in use by Athabasca University, which is a distance education institution and is overseeing this research.

1.3 Thesis Structure

Chapter 1 introduces the research, which is about personalizing a study guide for students based on the student's learning style. The learning style is predicted through a clustering analysis of student Web logs representing their sequential learning behaviours. The motivation and goal behind the research, to improve student learning, is addressed. Then, the purpose of the research, including the research question and hypothesis are discussed.

Chapter 2 describes the systematic literature review process that was undertaken. This chapter describes the 3 keyword searches that were done to cover the totality of the research scope. The chapter also indicates which journals were searched for, as well as how and why these journals were chosen among those available. The exclusion criteria used to eliminate less relevant articles is also addressed in the chapter.

Chapter 3 describes the Behaviour Analytics and LORD (Learning Object Relation Discovery) plugins including their underlying algorithms, inner workings, and performance. Behaviour Analytics is a plugin that uses interactive clustering to achieve student classification through sequential behaviour analysis. It builds a network graph from the learning objects in a course, where the nodes are positioned based on the relationship between the learning objects. The LORD plugin analyzes the relationship between learning objects and provides the node positioning. These plugins are detailed in the chapter.

Chapter 4 details the Personalized Study Guide plugin, which provides the main outcome of the research. The architecture and workflow of the plugin are presented. This plugin will require Behaviour

Analytics to provide the student classification, as this plugin is only responsible for the personalization of the learning objects within the study guide. This plugin provides the actual personalization aspect that the students will see.

Chapter 5 provides an evaluation of the research outcome. This includes the experiment design and the data analysis, as well as the findings and a discussion. Students will use the personalized study guide and then fill out a usability survey. Analysis of the survey results will provide the findings as to whether the plugin improves the learning outcome. The results will be discussed to explore their implications.

Chapter 6 concludes the thesis with a summary of the current research, as well as the limitations of the study and possible future works. The current research is limited in scope, not addressing all possible factors relating to it, and follows from the research of others, so may be the basis of future research by others. Such matters are discussed in the closing chapter.

Chapter 2. Systematic Literature Review (SLR)

The Systematic Literature Review process consists of choosing keywords, selecting journal sources, deciding the year range to search, finding articles, then further filtering the articles based on exclusion criteria. The keywords for this review consist of 3 sets of queries, as trying to combine all the keywords into a single query proved to overrun the search engine’s text entry capacity.

2.1 Existing Relevant SLRs

The SLR conducted for the current research focused on the 3 areas of sequence analysis, adaptive learning, and learning style prediction. Sequence analysis includes log file clustering, which is how the Behaviour Analytics plugin will group students. It is hoped that the clustering provided by Behaviour Analytics will coincide with the students' learning styles, thereby predicting learning style through a sequential analysis. The predicted learning style groups will be used to inform the personalization of the personalized study guide, providing adaptive learning to the students.

Vieira, Parsons, & Byrd (2018) explore visual learning analytics, which is at the intersection of learning analytics (analysis and reporting of student and learning data) and visual analytics (visual analysis and reporting of data). The goal of the analyzed studies is to provide visualizations of educational data, preferably novel and interactive. It is noted that most of the studies examined use simple visualizations such as bar, line, or pie charts and that the network graph is a popular visualization for social network analysis, but not used for other types of analysis. Vieira, Parsons, & Byrd (2018) categorize the studies regarding the purpose of the visualization, producing the 10 categories of understand collaboration, instructional design, understand a relationship, promote reflection, understand motivation, exploring usage patterns, explore learning progress, explore learning paths, understand discussion forums, and retention.

The Behaviour Analytics plugin of the current research could potentially be categorized under "understanding a relationship" (between learning behaviour and learning styles), "exploring usage patterns," or "exploring learning paths." This review calls for more innovative and interactive learning analytics tools to both enable personalized learning environments and to enable instructors to better understand the learning process. The Behaviour Analytics plugin is designed for visual learning analytics that is both innovative and interactive. The plugin will be used to provide personalized learning in the personalized study guide and may also help teachers to understand the learning process.

Martin et al. (2020) undertook a SLR examining adaptive learning, where 61 articles published between 2009 and 2018 were analyzed. Their review notes that there are 3 main models to consider in adaptive systems, which are the learner model, content model, and instructional model. The learner or student model includes data about the student characteristics such as attributes, preferences, and knowledge, where learning styles are the most common trait modeled, but learner behaviour is also considered. The content or domain model refers to the knowledge base of the course and includes the concepts and the relationships between them. The instructional or pedagogical model involves the algorithm that adapts the instruction based on the learner and content models, including sequencing of learning material. This review informs the current research as to how other adaptive systems are built and used, specifically that learning styles and student behaviour are common attributes in the student model, as are learning objects and their relationships in the content model, both of which will inform the instructional model when personalizing the study guide.

More specifically to student modeling is Afini Normadhi et al. (2019), where the authors focus on articles that identify learner traits. This review classifies learner traits into the 4 categories of cognition, affective, behaviour, and mixed, where the first 3 are from Bloom's taxonomy. Learning styles are included in the cognition category and are noted as being the most common trait considered in the reviewed studies. The detection of personal traits can be accomplished through a questionnaire, software, or a hybrid, where the questionnaire method is noted as having drawbacks that software detection can overcome. In the studies reviewed, most used software-based detection, but questionnaires are also used. One of the limitations noted is that the lack of data for validation of accuracy, reliability, and performance of trait identification, something the current research can assist with. Some supervised learning approaches also require large amounts of training data, while unsupervised clustering algorithms do not. The Behaviour Analytics plugin offers a novel clustering-based solution to detection of the learning style cognitive trait.

These 3 SLRs all covered different overlapping time spans, with Vieira, Parsons, & Byrd (2018) encompassing 17 years starting in 2000. Both Martin et al. (2020) and Afini Normadhi et al. (2019) reviewed a shorter time span, with Martin et al. (2020) covering 2009 to 2018, while Afini Normadhi et al. (2019) ranged from 2010 to 2017. The current SLR also overlaps this period heavily, but considers even more recent studies, covering 2011 to 2021. These SLRs also helped with keyword selection, as Vieira, Parsons, & Byrd (2018) dealt with visual analytics, which the current research does as well through visualization of sequential learning behaviour patterns captured in log files. The term personalized learning appears in the definition of adaptive learning taken by Martin et al. (2020), where these terms are used synonymously. Afini Normadhi et al. (2019) use a variety of words such as classify, detect, predict, identify, and estimate to describe the process of discovering learner traits.

2.2 SLR Process

The Behaviour Analytics plugin involves clustering of log entries representing the student's behaviour in the Learning Management System (LMS). So, the first set of keywords became (log OR sequence OR behaviour OR pattern) AND (cluster OR clustering), which formed the first query run in each of the journal search engines. The goal of clustering the student behaviour is to automatically detect the students' learning styles, so the second set of keywords used is "learning style" AND (predict OR detect OR identify OR discover OR estimate OR classify), which formed the second query run against the journal databases. Finally, the outcome of this research is a form of adaptive or personalized learning system, also called intelligent tutoring, so the final query used was (learning AND (adaptive OR adapt OR personalise OR personalised OR personalising)) OR (intelligent AND (tutor OR tutoring)).

To select journals to search, Google Scholar's Top Publication List (https://scholar.google.com/citations?view_op=top_venues) was used. From this list, the category "Engineering and Computer Science" was chosen, then the sub-categories of Education Technology, Data Mining and Analysis, and Artificial Intelligence were selected.

From the Education Technology sub-category, there are 20 publications available, as shown in Table 1.

TABLE 1: EDUCATION TECHNOLOGY PUBLICATIONS

	Publication	Relevant	Conference
1.	Computers & Education	X	
2.	British Journal of Educational Technology	X	
3.	The Internet and Higher Education		
4.	Journal of Educational Technology & Society	X	
5.	The International Review of Research in Open and Distributed Learning		
6.	Education and Information Technologies	X	
7.	Educational Technology Research and Development	X	
8.	International Conference on Learning Analytics & Knowledge		X
9.	Interactive Learning Environments		
10.	TechTrends	X	
11.	Journal of Computer Assisted Learning	X	
12.	Language Learning & Technology		
13.	Australasian Journal of Educational Technology	X	
14.	Computer Assisted Language Learning		
15.	International Journal of Educational Technology in Higher Education	X	
16.	IEEE Transactions on Learning Technologies	X	
17.	International Journal of Artificial Intelligence in Education	X	
18.	Learning, Media and Technology		
19.	TOJET: The Turkish Online Journal of Educational Technology	X	
20.	Journal of Educational Computing Research	X	

Of these, the International Conference on Learning Analytics & Knowledge is not a journal, so is the first to be excluded. There are also a number that are not as relevant to the current research, so these are also excluded, leaving 13 journals to search from the Education Technology sub-category. From the Data Mining and Analysis sub-category, there are 20 publications, as shown in Table 2.

TABLE 2: DATA MINING AND ANALYSIS PUBLICATIONS

	Publication	Relevant	Conference
1.	ACM SIGKDD International Conference on Knowledge Discovery and Data Mining		X
2.	IEEE Transactions on Knowledge and Data Engineering	X	
3.	International Conference on Artificial Intelligence and Statistics		X
4.	ACM International Conference on Web Search and Data Mining		X
5.	IEEE International Conference on Data Mining Workshop (ICDMW)		X
6.	ACM Conference on Recommender Systems		X
7.	Knowledge and Information Systems	X	
8.	IEEE International Conference on Big Data		X
9.	ACM Transactions on Intelligent Systems and Technology (TIST)		X
10.	Data Mining and Knowledge Discovery	X	
11.	Journal of Big Data	X	
12.	SIAM International Conference on Data Mining (SDM)		X
13.	European Conference on Machine Learning and Knowledge Discovery in Databases		X
14.	ACM Transactions on Knowledge Discovery from Data (TKDD)	X	
15.	Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery	X	
16.	Social Network Analysis and Mining	X	
17.	IEEE International Conference on Data Science and Advanced Analytics (DSAA)		X
18.	Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)		X
19.	Advances in Data Analysis and Classification	X	
20.	Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis		X

Of these, there are a number of conference publications which are excluded from further consideration, leaving 8 journals to search. From the Artificial Intelligence sub-category, there are another 20 publications, as shown in Table 3.

TABLE 3: ARTIFICIAL INTELLIGENCE PUBLICATIONS

	Publication	Relevant	Conference
1.	International Conference on Learning Representations		X
2.	Neural Information Processing Systems		
3.	International Conference on Machine Learning (ICML)		X
4.	AAAI Conference on Artificial Intelligence		X
5.	Expert Systems with Applications	X	
6.	IEEE Transactions On Systems, Man And Cybernetics Part B, Cybernetics	X	
7.	IEEE Transactions on Neural Networks and Learning Systems		
8.	Neurocomputing		
9.	Applied Soft Computing	X	
10.	International Joint Conference on Artificial Intelligence (IJCAI)		X
11.	IEEE Transactions on Fuzzy Systems		
12.	Knowledge-Based Systems	X	
13.	The Journal of Machine Learning Research	X	
14.	Neural Computing and Applications		
15.	Neural Networks		
16.	International Conference on Artificial Intelligence and Statistics		X
17.	Engineering Applications of Artificial Intelligence	X	
18.	Robotics and Autonomous Systems		
19.	Conference on Learning Theory (COLT)		X
20.	Journal of Intelligent & Fuzzy Systems	X	

Of these, there are a few conference publications, which are excluded from further consideration. There are also a few which are not as relevant to the current research, so are also excluded from further consideration, leaving 7 journals to search.

For the year publication range, only articles from the last 10 years (2011 - 2021) are considered for further evaluation. Searching the selected journals with the chosen keyword queries resulted in about 857 articles that seemed relevant, based on their titles, to which the exclusion criteria is applied to narrow down the final set. The target number of articles is between 40 and 60, so exclusion criteria are continuously applied until that number is met. These exclusion criteria included:

E-1. Reviews and meta-analysis

E-2. Does not involve log data or sequence analysis

E-3. Prediction of learning styles is not from student behaviour

E-4. Does not predict learning styles

E-5. Does not consider learning styles at all

E-6. Does not use a Learning Management System

After applying the exclusion criteria against the article abstracts, there were 60 articles remaining. The exclusion criteria were then applied against the entire text of the articles, further reducing the number to 44. The exclusion process is shown in Figure 1.

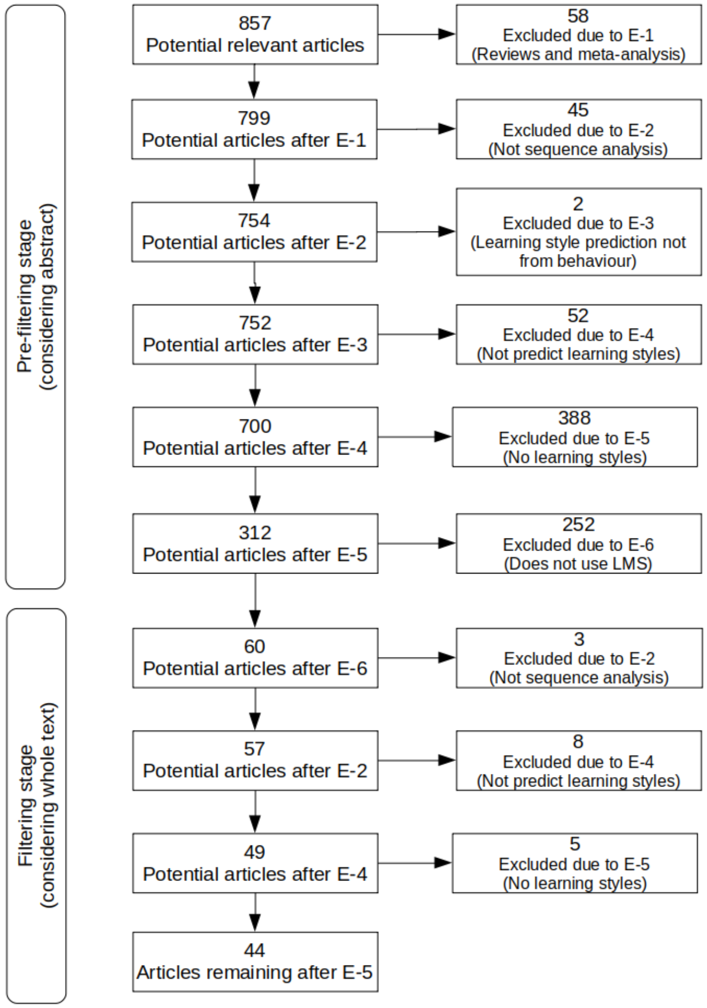


Figure 1: Literature Exclusion Process

2.3 Quantitative Analysis

To analyze the 44 articles found for this review, they can be grouped by topic. There are 3 distinct topics among these articles, each of which correspond to the keyword searches used. The first topic contains 30 articles and involves analysis of log file or behavioural sequence data. The current research aims to analyze student log file and behavioural sequence data with the Behaviour Analytics plugin. Every time a student accesses a learning object within the course, that access is logged, leaving a sequential trace of student behaviour that can be analyzed. The articles in this topic analyze similar sequential behaviour patterns in some way.

The second topic involves personalization of learning content and includes 19 articles. Personalized learning, adaptive learning, and intelligent tutoring all fall under the umbrella of the second topic. These systems offer different groups of learners different material or different layouts depending on the student's classification. The current research will also offer personalized learning determined by student learning style classification.

The final topic involves predicting learning styles and contains 16 articles. The current research offers personalized learning based on learning style classification, where student groups will be determined from their behaviour patterns embedded in their sequential learning object access data. Each group of students will receive a personalized study guide.

There is some overlap within these categories, as the personalization and prediction articles often use log or sequence data to achieve their goals, and some articles that predict learning styles also involve personalization. The above grouping includes such overlap, which is why there are more than the stated 44 articles. Figure 2 shows the article by topic count, while Table 4 lists which articles belong in which categories.

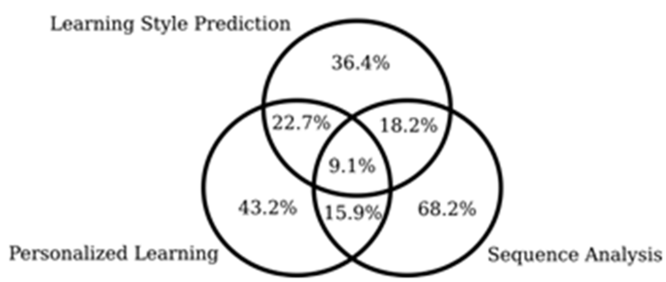


FIGURE 2: ARTICLES BY TOPIC

TABLE 4: ARTICLES BY TOPIC

Research	Sequence Analysis	Personalised Learning	Learning Style Prediction
Aeiad & Meziane (2019)		X	
Ahmad Uzir et al. (2020)	X		
Alian & Shaout (2017)			X
Ansari, Sattar, & Babu (2017)	X		
Antonenko, Toy, & Niederhauser (2012)	X		
Azzi et al. (2020)	X		X
Bernard et al. (2017)	X		X
Cabada et al. (2011)		X	X
Cerezo et al. (2016)	X		
Cinque et al. (2021)	X		
Dani & Nasser (2016)	X		
Dorça (2015)		X	X
Dorça et al. (2013)		X	X
Dwivedi, Kant, & Bharadwaj (2018)	X	X	X
El Aissaoui et al. (2019)	X		X
Ennouamani, Mahani, & Akharraz (2020)		X	
Fasihuddin, Skinner, & Athauda (2017)	X	X	X
Fatahi, Shabanali-Fami, & Moradi (2018)	X		X
Hoppe, Doberstein, & Hecking (2020)	X		
Huang et al. (2012)			X
Huang et al. (2016)	X		
Janning, Schatten, & Schmidt-Thieme (2016)	X		
Karagiannis & Satratzemi (2018)	X	X	X
Kassak, Kompan, & Bielikova (2016)	X		
Kim et al. (2018)	X		
Klašnja-Milićević et al. (2011)	X	X	X

Research	Sequence Analysis	Personalised Learning	Learning Style Prediction
Kolekar, Pai, & M. M. (2019)		X	X
Latham, Crockett, & McLean (2014)		X	X
Li & Tsai (2017)	X		
Li & Tsai (2020)	X		
Meza-Fernández & Sepúlveda-Sariego (2017)	X		
Niknam & Thulasiraman (2020)		X	
Poornalatha & Prakash (2013)	X		
Ramirez-Arellano et al. (2017)		X	
Rani, Nayak, & Vyas (2015)		X	X
Rastegarmoghadam & Ziarati (2017)		X	
Riestra-González, Paule-Ruíz, & Ortin (2021)	X		
Salehi & Nakhai Kamalabadi (2013)	X	X	
Salehi et al. (2014)	X	X	
Shi et al. (2020)		X	
Uto et al. (2020)	X		
Wan et al. (2012)	X		
Wan & Niu (2016)	X	X	
Xie, Zheng, & Zhang (2016)	X		

2.3.1 Sequence Analysis

The articles in the sequence analysis topic can be further analyzed by the sequential analysis techniques used, producing 6 categories, as listed in Table 5. These categories include those articles that apply basic clustering algorithms to the log file data directly, meaning that the log data is the input into the algorithm and the output classification is then used. There are also articles that use clustering as a secondary rather than primary method in the analysis, meaning that the log data is preprocessed by other algorithms, then the output of those algorithms is fed into the clustering algorithm.

Other forms of sequential behaviour analysis include that of association rule and sequential pattern mining, which are also included. There are a couple of articles that convert log file data to a network graph, which is what the Behaviour Analytics plugin does as well, so is included as a category as well. There are some articles that propose novel clustering algorithms, which is important to consider, as the utility of the algorithm can depend on the form of the data, and exploring different algorithms can help inform the decision as to the algorithm used in the current research. The final category is a collection of articles that conduct sequence analysis using other techniques such as neural networks and genetic algorithms.

TABLE 5: SEQUENCE ANALYSIS ARTICLES BY CATEGORY

Category	Articles	Percent (%)
Cluster log file directly	El Aissaoui et al. (2019), Azzi et al. (2020), Li & Tsai (2017), Ahmad Uzir et al. (2020), Kim et al. (2018), Janning, Schatten, & Schmidt-Thieme (2016), Cerezo et al. (2016), Li & Tsai (2020), Dani & Nasser (2016), Antonenko, Toy, & Niederhauser (2012)	33.3
Cluster as secondary analysis	Xie, Zheng, & Zhang (2016), Uto et al. (2020), Hoppe, Doberstein, & Hecking (2020), Poornalatha & Prakash (2013), Wan et al. (2012)	16.7
Association rules and sequential pattern mining	Wan & Niu (2016), Fatahi, Shabanali-Fami, & Moradi (2018), Klašnja-Milićević et al. (2011), Salehi & Nakhai Kamalabadi (2013), Salehi et al. (2014)	16.7
Other techniques	Karagiannis & Satratzemi (2018), Dwivedi, Kant, & Bharadwaj (2018), Bernard et al. (2017), Riestra-González, Paule-Ruíz, & Ortin (2021), Kassak, Kompan, & Bielikova (2016), Fasihuddin, Skinner, & Athauda (2017)	20.0
Convert log file to graph	Cinque et al. (2021), Meza-Fernández & Sepúlveda-Sariego (2017)	6.7
Propose new algorithm	Ansari, Sattar, & Babu (2017) ,Huang et al. (2016)	6.7

The sequence analysis articles can be analyzed based on the clustering algorithm used, for those that use clustering with log data. Several different clustering algorithms are used, as listed in Table 6. Because not all the articles use clustering, some articles are not represented in Table 6. It is also the case that some researchers used multiple clustering algorithms, either to validate or compare results from other algorithms. It can be seen in Table 6 that the k-means algorithm dominates the selected literature, with hierarchical clustering coming in a distant second. The k-means algorithm is very common, but perhaps this is only because it is simple to implement.

TABLE 6: CLUSTERING ALGORITHMS USED IN SEQUENCE ANALYSIS ARTICLES

Algorithm	Articles	Percent (%)
K-means	El Aissaoui et al. (2019), Wan & Niu (2016), Li & Tsai (2017), Salehi & Nakhai Kamalabadi (2013), Cerezo et al. (2016), Li & Tsai (2020), Antonenko, Toy, & Niederhauser (2012), Poornalatha & Prakash (2013), Wan et al. (2012)	30.0
Hierarchical	Li & Tsai (2017), Ahmad Uzir et al. (2020), Antonenko, Toy, & Niederhauser (2012)	10.0
C-means	El Aissaoui et al. (2019), Azzi et al. (2020)	6.7
Neural Network	Ansari, Sattar, & Babu (2017), Bernard et al. (2017)	6.7
Expectation Maximization	Cerezo et al. (2016)	3.3
K-Nearest Neighbours	Janning, Schatten, & Schmidt-Thieme (2016)	3.3
Partitioning Around Medoids	Hoppe, Doberstein, & Hecking (2020)	3.3
K-medoids	Kim et al. (2018)	3.3
Spectral	Xie, Zheng, & Zhang (2016)	3.3
Log-likelihood	Dani & Nasser (2016)	3.3
Multiple non-negative matrices factorization	Huang et al. (2016)	3.3
Unspecified	Cinque et al. (2021), Klašnja-Milićević et al. (2011), Uto et al. (2020)	10.0

2.3.2 Personalized Learning

The 19 articles involving personalized learning can be analyzed based on whether they consider learning styles in the personalization and by the techniques and algorithms used. Of the 19 articles, 15 consider the student's learning style when providing personalization, while the other 4 do not. Of the 15 articles that consider learning style, most of them utilize the Felder-Silverman Learning Style Model (FSLSM).

There are many models, theories, and related terms such as learning strategy, learning preference, and cognitive style. Within learning style research there are authors who make clear distinctions between the various terms and others who do not (Cassidy, 2010).

The FSLSM was specifically designed for use in the engineering disciplines. Originally it considered 5 dimensions in the way students learn related to perception, input, organization, processing, and understanding.

- Perception dimension - Considers sensing versus intuitive, where sensing learners prefer taking in information through the senses while intuitive learners prefer indirectly perceiving information through the unconscious by way of speculation and imagination.
- Input dimension - Has visual as one pole and verbal as the other, where visual learners prefer videos and images to written text or spoken language.
- Organization dimension - Considers inductive versus deductive, where inductive learners prefer to induce principles from concrete observations and deductive learners prefer deducing the consequences given the underlying principles.
- Processing dimension - Considers active versus reflective, where active learners prefer experimenting in the world with their information and reflective learners prefer examining or manipulating their information introspectively.
- Understanding dimension - Has sequential at one pole and global at the other, where sequential learners prefer material to be logically ordered by time and global learners prefer material in less organized way as they do not learn in a steady or predictable manner.

After much testing, the organization dimension was removed, leaving the other 4. Each dimension is measured on a scale from 0 to 11, where the higher the number the stronger the preference. This means that a learner might be classified as global/verbal/sensing/active or global/visual/intuitive/active. Using such classification results in 16 possible learning styles, but the classification can be strong or weak. For instance, a learner may score a 1 or a 3 for a given dimension making the classification weak. But a learner may also score 9 or 11, which indicates a strong preference for that dimension. Across the 4 dimensions of the classification some dimensions may show strong preference and others weak.

Learning style is determined by way of a questionnaire called the Index of Learning Styles. It consists of 44 questions with binary answers where there are 11 questions focus on each dimension. The final score for a dimension is determined by subtracting the results of the poles. For example, if a learner scored 6 for global and 5 for sequential, then the learner is classified as global with a score of 1. Likewise, if the learner scored a 10 for verbal and 1 for visual they would be classified as verbal with a score of 9 (Felder & Silverman, 1988).

Another popular learning style model is the Visual, Auditory, Reading/Writing, Kinesthetic (VARK) model (VARK, n.d.). The VARK model shares with the FSLSM the consideration of a visual modality, but considers spoken and written language to be separate modalities. The VARK model also considers a kinesthetic modality, which are hands-on learners that like to learn by doing (Fleming, 1995). To determine a VARK score, a learner takes a questionnaire with 16 four answer questions where the learner may select multiple answers for each question. The result is a set of scores for each modality, which is then translated into categories. There are 25 different categories in the model which are the single preferences: very strong, strong, and mild Visual, Aural, Read/write and Kinesthetic, and also the multimodal combinations: VARK Type One, VARK Type Two, VARK Transition, VAR, VAK, VRK, ARK, VA, VR, VK, AR, AK, and RK.

Another widely used model is the Myers-Briggs Type Indicator (MBTI) (The Myers-Briggs Company, n.d.), which shares with the FSLSM grading learners in 4 dimensions and even shares the Sensing/iNtuition (S/N) dimension. The other 3 dimensions of the model are Extroversion/Introversion (E/I), Thinking/Feeling (T/F), and Judging/Perceiving (J/P). The E/I dimension considers the opposite ways people direct and receive energy, which are outward into the world or inward into themselves. The S/N dimension considers the opposite ways people take in information, which are from the world through 5 senses or from associations and possibilities. The T/F dimension considers the opposite ways people make decisions, which are dispassionately with objective logic or based on human values and motives. The J/P dimension considers the opposite ways people approach the outside world, which are planned and organized or spontaneous and adaptable. The result of the questionnaire is one of 16 possible 4 letter classifications such as ENTP or ISFJ (https://www.themyersbriggs.com/en-US/Products-and-Services/Myers-Briggs).

There are a variety of common algorithms and techniques such as Genetic Algorithm and Natural Language Processing (NLP) used in these articles, the most common of which were association rule mining and clustering, both of which made 4 appearances in the selected articles, and Ant Colony Optimization, which was seen 2 times. Only 1 article mentions SCORM compliance or other standard learning object classification. There is 1 article that deals with a conversational intelligent tutoring and 2 that are specifically text query driven. Most important to the present research are the articles related to personalization based on learning styles. Table 7 summarizes the types of LS model used and that the FSLSM is the most widely used learning style model. Each LS model has its own questionnaire to determine a student's learning style such as the FSLSM that uses the Index of Learning Styles.

TABLE 7: LEARNING STYLE MODELS USED IN PERSONALIZED LEARNING

	Articles	Percent (%)
FSLSM	Ennouamani, Mahani, & Akharraz (2020), Latham, Crockett, & McLean (2014), Karagiannis & Satratzemi (2018), Rani, Nayak, & Vyas (2015), Klašnja-Milićević et al. (2011), Cabada et al. (2011), Dwivedi, Kant, & Bharadwaj (2018), Kolekar, Pai, & M. M. (2019), Fasihuddin, Skinner, & Athauda (2017)	69.2
VARK	Aeiad & Meziane (2019), Rastegarmoghadam & Ziarati (2017)	15.4
Other	Wan & Niu (2016), Ramirez-Arellano et al. (2017)	15.4

2.3.3 Learning Style Prediction

The 16 articles dealing with learning style prediction can be analyzed by which learning style (LS) model is used. These articles, like those in the personalization topic, are dominated by the Felder-Silverman (FS) model, with 12 of the 16 articles (75%). These articles can also be analyzed by the techniques used in the prediction and the number of learning style groups considered, Table 8 lists the various techniques used, where they were used, which LS model is used, and how many LS groups are considered. The FSLSM has 4 main dimensions that can be considered as perception, input, processing, and understanding, where each dimension has 2 directions measured on a scale, which gives rise to some researchers considering 4 groups (main dimensions), some 8 (2 directions for each of 4 dimensions), and others 16 (all possible combinations or 24). Of the articles that use neural networks, Bernard et al. (2017) utilizes supervised learning with behaviour patterns, while Cabada et al. (2011) employ a self-organizing map, but considers performance rather than student behaviour.

TABLE 8: LS PREDICTION ARTICLES BY TECHNIQUE WITH MODEL AND LS GROUPS

Technique	Article(s)	LS model	Groups
Log file clustering	El Aissaoui et al. (2019), Azzi et al. (2020), Kolekar, Pai, & M. M. (2019)	FS, FS, FS	16, 8, 8
Rule based	Alian & Shaout (2017), Huang et al. (2012), Rani, Nayak, & Vyas (2015), Fasihuddin, Skinner, & Athauda (2017)	VARK, FS, FS, FS	4, 4, 4, 4
Sequential pattern mining	Fatahi, Shabanali-Fami, & Moradi (2018), Klašnja-Milićević et al. (2011)	MBTI, FS	4, 16
Decision tree	Karagiannis & Satratzemi (2018)	FS	4
Neural network	Cabada et al. (2011), Bernard et al. (2017)	FS, FS	4, 8
Reinforcement learning	Dorça et al. (2013), Dorça (2015)	variable, variable	4, 4
Genetic algorithm	Dwivedi, Kant, & Bharadwaj (2018)	FS	4

The accuracy and precision of the learning style prediction can also be analyzed, summarized in Table 9. Some of the prediction occurs as part of personalization systems, so is not always measured directly and the personalization is measured instead. Only articles that measure the prediction are considered in Table 9. All but one of the articles lists accuracy or precision results as measured against the questionnaire, which makes it easy to compare these articles to each other and to the current research, while the other article conducts usability testing.

TABLE 9: METHOD, MEASUREMENT, AND RESULTS OF LS PREDICTION ARTICLES

Article	Method	Measures	Results
El Aissaoui et al. (2019)	comparison with questionnaire	accuracy, precision, recall, F1	A=96.89%, P=97.27% R=99.44%, F1=98.53%
Azzi et al. (2020)	comparison with questionnaire	accuracy	Test1=93.41%, Test2=93.16%
Alian & Shaout (2017)	comparison with questionnaire	accuracy	48.5%
Fasihuddin, Skinner, & Athauda (2017)	comparison with questionnaire	accuracy	Perception 71.23%, Input 80.13%, Processing 78.08%, Understanding 81%
Fatahi, Shabanali-Fami, & Moradi (2018)	comparison with questionnaire	accuracy	Introversion/Extroversion 78.5%, Feeling/Thinking 79.25%, Intuition/Sensing 77.25%, Judging/Perceiving 77.75%
Karagiannis & Satratzemi (2018)	comparison with questionnaire	precision	Active/Reflective 70%, Sensing/Intuitive 66%, Visual/Verbal 75%, Sequential/Global 80%
Bernard et al. (2017)	comparison with questionnaire	accuracy	80.7%
Rani, Nayak, & Vyas (2015)	custom questionnaire	scale of 1-5	Learner 3.87, Instructor 3.45, Course 3.74, Technology 3.45, Design 3.659

There are some articles that use fuzzy classification when predicting learning style, summarized in Table 10. Of these, both El Aissaoui et al. (2019) and Azzi et al. (2020) use Fuzzy C-Means (FCM), but only El Aissaoui et al. (2019) apply defuzzification after convergence to produce a single learning style for the student. The word defuzzification is not mentioned by Azzi et al. (2020), which reports highly accurate results, but does not indicate whether the results include the fuzzy memberships. It can be seen in Table 9 that the results by Alian & Shaout (2017) are poor, but above 90% by both El Aissaoui et al. (2019) and Azzi et al. (2020), where El Aissaoui et al. (2019) scored higher than Azzi et al. (2020) despite using defuzzification. The FCM clustering algorithm was directly compared to the k-means by El Aissaoui et al. (2019), reporting that FCM is the superior of the 2, and also that deriving a single classification from a fuzzy one can allow for the inclusion of both types of clustering. Both El Aissaoui et al. (2019) and Azzi et al. (2020) use log file clustering to produce superior prediction results, indicating that behaviour sequences stored in the log file are useful for predicting learning styles.

Table 10: Articles Using Fuzzy Classification and Defuzzification

Article	Technique	Defuzzification
El Aissaoui et al. (2019)	Fuzzy C-Means	Maximal value of membership
Azzi et al. (2020)	Fuzzy C-Means	None
Alian & Shaout (2017)	Rule based	Centroid technique

2.4 Qualitative Analysis

2.4.1 Learning Objects

There are 17 articles analyzed that consider Learning Objects (LO) in some way. A learning object is:

> "... an elementary unit of learning which construct a learning content. The LO can take various digital resources forms as text, image, video, audio, etc. and it is characterized by its ability to be reusable, interoperable, durable, and accessible." (El Aissaoui et al., 2019)

There are initiatives such as the Shareable Content Object Reference Model (SCORM) (Advanced Distributed Learning Initiative, n.d.) that offer standards to follow for LO interoperability and sequencing. Unfortunately, of the 17 articles, only Ramirez-Arellano et al. (2017) consider SCORM in their research. This makes comparing the various studies more difficult, as most adopt their own definitions about learning objects and their attributes. Shi et al. (2020), for instance, consider only the 3 LO types of basic knowledge, algorithm, and task, which may not be applicable in other contexts.

There are a few articles that consider learning objects when predicting Learning Styles (LS) by mapping the LO to a dimension of the learning style model. Such was the case by El Aissaoui et al. (2019), Azzi et al. (2020), and Bernard et al. (2017), where the prediction was the only goal of the researchers. Learning style prediction is a common practice with personalized and adapted learning systems, where the LO are also mapped to LS dimensions, as was done by Karagiannis & Satratzemi (2018), Dorça et al. (2013), Cabada et al. (2011), Dorça (2015), Kolekar, Pai, & M. M. (2019), and Fasihuddin, Skinner, & Athauda (2017). There are also personalized learning systems that map LO to LS, but do not use any automatic LS prediction, relying instead on questionnaires, as by Ennouamani, Mahani, & Akharraz (2020) and Ramirez-Arellano et al. (2017).

Tables 11 through 14 show some examples of mapping learning object to learning style. The data in Table 11 shows a LO to LS preference mapping which contains a subset of the data presented in Table 3 from Ramirez-Arellano et al. (2017). The data in Table 12 is an abbreviated version of Table 2 by El Aissaoui et al. (2019), which considers all 16 possible LS dimensions of the FSLSM, each as a different cluster. The data in Table 13 has been taken from Table 1 by Azzi et al. (2020) and the data in Table 14 comes from Table 5 by Kolekar, Pai, & M. M. (2019), where both consider 8 dimensions of the FSLSM and have almost identical LO to LS mappings. While these last 2 tables are more easily comparable with each other than the first 2 tables, there are some discrepancies between the mappings, as Azzi et al. (2020) consider videos to be both visual and verbal, while Kolekar, Pai, & M. M. (2019) list videos as visual, but not verbal. The variation among mappings make comparison difficult and indicates that care must be taken to construct a correct mapping to have meaningful results.

TABLE 11: FROM TABLE 3. MATRIX FOR MAPPING LOM VALUES TO PREFERENCES RAMIREZ-ARELLANO ET AL. (2017)

	Perception		Processing information				Reasoning	
	Visual	Verbal	Abstract	Concrete	Active	Reflective	Deductive	Inductive
Video	X							
Lecture		X					X	X
Expositive			X					
Simulation					X		X	X
Exercise					X			
Questionnaire						X		
Problems							X	X

TABLE 12: FROM TABLE 2 LEARNING OBJECTS AS PER FSLSM EL AISSAOUI ET AL. (2019)

ClusterID	Combination	Videos	PPTs	Demo	Assignments	Forum	Images	Charts	Email
C01	(R, I, Ve. G)	3	2	0	1	1	1	1	0
C02	(A, I, Ve, G)	3	2	1	2	1	1	1	0
C03	(R, Sen, Ve, G)	3	1	0	1	0	1	1	0
...C15	(R, Sen, Vi, Seq)	3	1	0	1	0	1	1	0
C16	(A, Sen, Vi, Seq)	3	1	1	2	0	1	1	0

TABLE 13: FROM TABLE 1 MAPPING OF LEARNING OBJECTS AS PER FSLSM AZZI ET AL. (2020)

FSLSM Dimensions	Learning Objects
Active	Videos, PPTs, Demo, Exercise, Assignments
Reflective	PDFs, PPTs, Videos, Announcements, References
Sensing	Examples, PDFs, Videos, Practical Material
Intuitive	PDFs, PPTs, Videos, Forum, Topic, List, References
Visual	Images, Charts, Videos, References
Verbal	PDFs, Videos, Email, Announcements
Sequential	Exercise, References, Assignments, Sequential
Global	Topic Lists, References, Exercise, Assignment

TABLE 14: FROM TABLE 5 LEARNING OBJECTS MAPPED TO FSLSM CATEGORIES KOLEKAR, PAI, & M. M. (2019)

Learning styles	Learning objects
Active	Videos, PPTs, demos, exercise, assignments, forum, announcements
Reflective	PDFs, PPTs, videos, announcements, references, email
Sensing	Examples, PDFs, videos, practical material, forum
Intuitive	PDFs, PPTs, videos, forum, TopicList, references, assignments, advanced topics
Visual	Images, charts, videos, references
Verbal	PDFs, email, announcements
Sequential	Exercise, references, assignments, sequential links, email
Global	Topic lists, references, exercise, assignments, PPTs, forums

Learning paths are a common approach to personalized learning, as was done by Wan & Niu (2016), Shi et al. (2020), Cabada et al. (2011), Rastegarmoghadam & Ziarati (2017), Dwivedi, Kant, & Bharadwaj (2018), and Niknam & Thulasiraman (2020), where a learning path is comprised of a sequence of learning objects that is recommended to the student. In the system described by Dwivedi, Kant, & Bharadwaj (2018), ratings of the learning object by other learners are considered when building the learning path. Knowledge is also considered in some personalization systems, as is the case by Wan & Niu (2016) and Niknam & Thulasiraman (2020), where the learning objects can have a difficulty rating and the learner has a corresponding level of knowledge to further aid the learning path recommendation.

2.4.2 Sequence Analysis

There are 30 articles in this topic that can be analyzed according to the methods they use to conduct the sequential analysis. Clustering is used in the Behaviour Analytics plugin, so articles involving clustering are of greater significance, and a variety of clustering algorithms have been used in various ways. There are a number of articles where the clustering was a secondary analysis method rather than the primary method and these articles are treated as forming a distinct group from those where clustering is the lone or primary analysis method and those that present novel clustering algorithms. However, clustering is not the only way to classify students from their sequence data, so other techniques are considered separately. Also, classification is also not the only goal of sequence analysis, so association rule mining and sequential pattern analysis are also discussed. Finally, there are a couple of articles dealing with converting log data to network graphs for analysis, which is what Behaviour Analytics does, so these articles are discussed together.

Many of the articles in the log data and sequence analysis category use basic clustering directly on the log file to analyze student or user behaviour. Such is the case with Li & Tsai (2017), where hierarchical clustering is used to determine the optimal number of clusters, then k-means is used to discover 3 different groups of student behaviour patterns. K-means is used again by Li & Tsai (2020) to discover 3 different groups of student's patterns of accessing time of learning resources. There were four time management strategies used by students identified using hierarchical clustering by Ahmad Uzir et al. (2020). Cerezo et al. (2016) also found four clusters of students using expectation maximization clustering, which they confirmed using k-means.

The k-medoids algorithm was used by Kim et al. (2018) to discover three self-regulated learning profiles. The log-likelihood algorithm was utilized by Dani & Nasser (2016), where the researchers determined that three different groups of students existed based on the number of topics studied to the number of topics mastered in the course, with a correlation to final marks. Both hierarchical and k-means clustering were used by Antonenko, Toy, & Niederhauser (2012) to analyze four groups of learners with regard to their problem solving activity. The k-nearest neighbours approach was used by Janning, Schatten, & Schmidt-Thieme (2016), along with other machine learning algorithms for supporting perceived task-difficulty recognition and supporting task sequencing. Both El Aissaoui et al. (2019) and Azzi et al. (2020) apply fuzzy c-means to student log data to predict learning styles, where El

Aissaoui et al. (2019) compares the c-means results to the k-means algorithm and Azzi et al. (2020) consider time spent on the learning resource in addition to its access.

Some of the articles use clustering as a secondary method in the analysis rather than as a primary method. Although Uto et al. (2020) utilize clustering and log data, it does not use typical Web log data. Because Uto et al. (2020) analyze student writing, the log data is generated from students' typing of characters. The analysis by Uto et al. (2020) first applies a Gaussian Hidden Markov Model, then a secondary analysis uses an unspecified clustering algorithm to group students into different writing behaviour patterns. The analysis by Hoppe, Doberstein, & Hecking (2020) also used secondary clustering after using sequence analysis between sets of log data to produce a similarity matrix, which was the basis of a partitioning around medoids (PAM) algorithm.

A Hybrid Sequence Alignment Measure (HSAM) was used by Poornalatha & Prakash (2013) as a primary analysis of user sessions of unequal length, then a modified k-means algorithm was used for clustering as a secondary analysis. K-means was also used as a secondary analysis by Wan et al. (2012) after a random indexing scheme was used on Web users' transaction data. Xie, Zheng, & Zhang (2016) started by converting log sequences into trajectories, then used the similarities between trajectories to construct a similarity matrix and a graph, and finally partitioned the graph using spectral clustering.

Of the articles that do not use a clustering approach, both Riestra-González, Paule-Ruíz, & Ortin (2021) and Kassak, Kompan, & Bielikova (2016) use other machine learning approaches for Web session exit prediction (Kassak, Kompan, & Bielikova, 2016) and for course failure prediction (Riestra-González, Paule-Ruíz, & Ortin, 2021). The system by Dwivedi, Kant, & Bharadwaj (2018) uses a variable length genetic algorithm and historical student data to personalize the learning path for active students. Bernard et al. (2017) used the genetic algorithm, neural network, particle swarm, and ant colony system are compared when predicting learning styles from behaviour patterns. Both Karagiannis & Satratzemi

(2018) and Fasihuddin, Skinner, & Athauda (2017) compare predetermined behaviour patterns with the actual student behaviour for the purpose of learning style prediction.

Other common sequence analysis techniques are that of association rule and sequential pattern mining. Wan & Niu (2016) mine association rules from student log data to find the relationships between learning styles and learning objects, then k-means clustering is applied. The system by Klašnja-Milićević et al. (2011) first uses clustering to determine learning style categories, then uses AprioriAll to mine association rules for the purpose of personalization. The systems in both Salehi & Nakhai Kamalabadi (2013) and Salehi et al. (2014) mine association rules with the Apriori algorithm and make use of PrefixSpan for sequential pattern mining to build a learner preference tree for personalization, where the difference is that Salehi & Nakhai Kamalabadi (2013) utilize k-means clustering and Salehi et al. (2014) uses collaborative filtering. The system described by Fatahi, Shabanali-Fami, & Moradi (2018) makes use of sequential pattern mining to predict students' learning styles.

The systems by Cinque et al. (2021) and Meza-Fernández & Sepúlveda-Sariego (2017) are interesting primarily because they convert the log file data into a network graph for further analysis, which is similar to the approach used by the Behaviour Analytics plugin. The system by Meza-Fernández & Sepúlveda-Sariego (2017) attempts to classify students into learning style categories by their behaviour patterns embedded in the student's Web browsing activity. Cinque et al. (2021) first used clustering on the log entries to group similar behaviours, then templates are derived from the clusters, which are used to build the graphs. Unlike Meza-Fernández & Sepúlveda-Sariego (2017), however, the goal of Cinque et al. (2021) was to detect anomalous user behaviour using path finding rather than analyze different patterns of behaviour.

There are 2 articles that present new algorithms for clustering sequence data. A fuzzy neural network is proposed by Ansari, Sattar, & Babu (2017) that combines a modified self-organizing maps and fuzzy c-means to cluster Web access data. Huang et al. (2016) introduce a Multiple Non-negative Matrices Factorization method to cluster time stamped data and discover evolutionary trends.

Both Cinque et al. (2021) and Meza-Fernández & Sepúlveda-Sariego (2017) share the idea of building a network graph from the log file data, but conduct a different analysis on the resulting graph than Behaviour Analytics does. Many of the included articles cluster student behaviour from the log data in some way, showing that there are many ways to cluster sequence data, but none provide an interactive visual framework to complement the cluster analysis. Some of the other articles involve clustering as a secondary analysis method, indicating that clustering does not have to be applied directly to the log data to provide meaningful classification. The k-means algorithm is the most common clustering algorithm, but this may be due solely to its simplicity. Other means of classification and analysis are also valid approaches that could be used to confirm the results achieved with clustering.

The current research, through the Behaviour Analytics plugin, intends to provide an interactive visual learning analytics tool that can cluster students based on their sequential behaviour as stored in a log file. Behaviour Analytics uses clustering as a secondary analysis method after converting the log file into a network graph and compressing the student behaviour into a centroid point. The visual framework provides coordinate points for the centroids and lends itself to a simple Euclidean distance function, as all clustering algorithms require a distance function. The k-means algorithm operates over multiple iterations, allowing the clustering to be observed and interacted with, making k-means a good choice for use in the Behaviour Analytics plugin.

2.4.3 Personalized Learning

The second topic involves personalization of learning content and includes 19 articles, which can be analyzed according to how they make use of different learning style models, where the use of the Felder-Silverman model (FSLSM) and its Index of Learning Styles (ILS) questionnaire are the most common. The articles can also be analyzed with respect to the various algorithms used, techniques applied, and personalization approaches used. Some of the personalization approaches stem from learning styles, while others do not. Some of the techniques are behaviour based, while others are not.

Of the articles that consider learning styles, all but Kolekar, Pai, & M. M. (2019) have students fill in a questionnaire as a baseline measure. Of these, Wan & Niu (2016) define their own questionnaire and Ramirez-Arellano et al. (2017) use the Wayne State University Learning Styles Questionnaire (WSULSQ), both of which are based in part on the FSLSM. Both Aeiad & Meziane (2019) and Rastegarmoghadam & Ziarati (2017) use the Visual-Aural-Read/Write-Kinesthetic (VARK) model, while the rest rely on the Felder-Silverman model and the ILS questionnaire. Kolekar, Pai, & M. M. (2019) also used the FSLSM, but did not administer a questionnaire to students, and determined the student's learning style from monitoring their resource accesses.

Monitoring student behaviour and using log data to determine, validate, and update learning style preferences is a common practice among recommendation systems. Both Wan & Niu (2016) and Klašnja-Milićević et al. (2011) apply association rule mining to student access data to find relationships between learning objects accessed and learning styles. The system by Klašnja-Milićević et al. (2011) may also alter the student's learning style based on their behaviour. Kolekar, Pai, & M. M. (2019) monitor behaviour and define a rule based system that makes recommendations based on observed behaviour and can also adapt a student's style if their behaviour changes. The system by Dwivedi, Kant, & Bharadwaj (2018) uses student logs and a variable length genetic algorithm to update their learning path recommendation as they progress through a course. Rani, Nayak, & Vyas (2015) monitor the behaviour of students to adapt their learning style category if the accessed resources differ from their initial questionnaire assigned category. Both Karagiannis & Satratzemi (2018) and Fasihuddin, Skinner, & Athauda (2017) give students the ILS to determine personalized course material presentation, then the students' access patterns are monitored to determine if their access patterns are in line with the questionnaire results. Both Dorça et al. (2013) and Dorça (2015) also base material presentation on learning style and continuously adjust the presentation using simulated data, but with different learning algorithms.

Other learning style, but not behaviour based, approaches to personalization include Ennouamani, Mahani, & Akharraz (2020), where ILS questionnaire results are used to determine the format of the learning material, either text or video, and the menu style, either sequential or global. The system by Ennouamani, Mahani, & Akharraz (2020) is not behaviour driven nor is it capable of altering a student's style after the questionnaire, but the authors report impressive gains from the students using the system. The Conversational Intelligent Tutoring System (CITS) used by Latham, Crockett, & McLean (2014) uses the FSLSM as well, but rather than presenting learning objects to students, the tutorial conversation is directed to accommodate different learning styles. Both Aeiad & Meziane (2019) and Rastegarmoghadam & Ziarati (2017) use the VARK model, but where Rastegarmoghadam & Ziarati (2017) improved an existing system, Aeiad & Meziane (2019) use Natural Language Processing to select learning content from the WWW and structure a course with it according to a standard curriculum, similar to what a human tutor would do. The system described by Ramirez-Arellano et al. (2017) is the only one to use the WSULSQ and also the only article to merge learning objects together to create better ones, which it does by using an encoded genealogical tree algorithm. The system by Cabada et al. (2011) inputs the students learning style, learning objects associated style, and student marks into a self-organizing map to recommend learning objects, where the student's style may change after taking a quiz.

Behaviour based approaches that do not consider learning style include both Salehi & Nakhai Kamalabadi (2013) and Salehi et al. (2014). Both articles share authors, similar publication years, and introduce a Learner Preference Tree (LPT). Both articles describe the use of sequential- and attribute-based recommendation, where the sequential aspect derives from student's sequential learning material access patterns.

Other approaches include those mentioned by Shi et al. (2020) and Niknam & Thulasiraman (2020), both of which consider knowledge in their approach. The system by Shi et al. (2020) uses a knowledge graph to recommend three different kinds of learning objects, where the output learning path is comprised of knowledge, algorithm, and task type material, and is text query driven. Niknam & Thulasiraman (2020) cluster on the previous knowledge of the student, then use ant colony optimization to select the optimal learning path from the cluster for the recommendation.

Like many of the articles discussed in this topic, the personalization in the current research will be based on a learning style model, of which the FSLSM is the most used. The Behaviour Analytics plugin involved in the current research is behaviour based and will monitor student behaviour to adapt the learner model and recommendation accordingly. There are many ways to make recommendations to students without monitoring their behaviour, and such techniques could also be used to augment a behaviour-based recommendation approach. Likewise, consideration of learning styles is not necessary and other approaches could be fused with learning styles to further enhance the personalization.

2.4.4 Learning Style Prediction

The final topic of predicting learning styles contains 16 articles, which can be analyzed according to the type of learning style model used, where the FSLSM dominates again. The articles can also be analyzed with respect to the techniques utilized, algorithms applied, and how the prediction is used. Some systems are rule based, some use crisp classification, and others use fuzzy techniques. Also, the learning style prediction can be the output of the system or used as part of a larger personalization system.

The learning style models used in these articles is dominated by the Felder-Silverman Learning Style Model (FSLSM), but some other models are used as well. Both Dorça et al. (2013) and Dorça (2015) use simulated data and are set up to incorporate any model its users prefer, while Fatahi, Shabanali-

Fami, & Moradi (2018) use the Meyers-Briggs Type Indicators (MBTI), and Alian & Shaout (2017) use the Visual-Aural-Read/Write-Kinesthetic (VARK) model.

Some of the articles in this topic include rule-based systems, such as the work by Alian & Shaout (2017), in which the authors feed the results of two student questionnaires, VARK and one of their own, into the system to output learning style. The system by Kolekar, Pai, & M. M. (2019) uses rule-based personalization, where the student behaviour triggers rules about which learning style dimension they fit to inform the adaptation. The system by Huang et al. (2012) uses Cat Optimization and allows the user to define their own rules that determine which student actions belong to which learning style dimension.

There are some articles that consider a fuzzy classification of learners into multiple dimensions rather than into a single category. The systems described in both El Aissaoui et al. (2019) and Azzi et al. (2020) use fuzzy c-means to mine learning style groups from the student access logs. Both Alian & Shaout (2017) and Huang et al. (2012) describe rule-based systems, but also take a fuzzy approach. In Alian & Shaout (2017) a student may have 2 or even 3 learning style preferences, while Huang et al. (2012) use Cat Optimization-based fuzzy knowledge fusion with user supplied rules. Huang et al. (2012) apply fuzzy optimization to the rules rather than to the learning style classification.

Learning style prediction can be done by itself, as described by Fasihuddin, Skinner, & Athauda (2017), where students fill out the Index of Learning Styles (ILS) questionnaire and then the results are compared against the student access patterns to gauge the accuracy of the questionnaire. Prediction was the only goal by Bernard et al. (2017) as well, where an artificial neural network was deemed the best approach against a genetic algorithm, particle swarm optimization, and an ant colony system. The system described by Karagiannis & Satratzemi (2018) also compares ILS results with access patterns but includes a personalization aspect as well.

Some of the learning style prediction is based on adapting the personalization to changing learning styles. Such is the case with Rani, Nayak, & Vyas (2015), where the system presents material according to the student's preferred style as detected through the ILS, but monitors student behaviour and can update the student model as they progress through the course. The system proposed by Klašnja-Milićević et al. (2011) also monitors student access patterns and updates the student model when their patterns match that of a different cluster of learners. Dwivedi, Kant, & Bharadwaj (2018) initially presented learners with a predefined learning path that is personalized as the student progresses, requiring that the student model be updated continuously. The system by Cabada et al. (2011) can update the student's learning preferences any time the student takes a quiz, which tests their knowledge, given the form of materials they were presented. The system by Latham, Crockett, & McLean (2014) updates the student model according to conversational cues given by the learner during a conversational tutorial.

The current research aims to predict learning styles as part of a personalization system rather than as the goal of system. The system may apply simple rules to determine the personalization, but is not a rule based system in the same sense as those described above. The Behaviour Analytics plugin uses regular clustering rather than fuzzy, but the fuzzy approach is still valid and may be worth considering in the future.

The current research will, like many other studies, use the FSLSM (Felder & Silverman, 1988). The Felder-Silverman model is reported to be the most appropriate model for use in personalized learning (El Aissaoui et al., 2019; Bernard et al., 2017; Fasihuddin, Skinner, & Athauda, 2017) and also has a reliable questionnaire (Bernard et al., 2017; Fasihuddin, Skinner, & Athauda, 2017), as well as being both comprehensive and simple, with measurable dimensions that readily map to e-learning aspects (Rani, Nayak, & Vyas, 2015).

2.5 Research questions

The purpose of the current research is to automatically detect learning styles from student behaviour and provide students with an enhanced personalized learning experience. To achieve this, the study guide layout will be personalized to each student based on their learning style, which will be detected through the use of the Behaviour Analytics plugin. However, the data collected determine the research questions that can be asked, a problem detailed in Sections 6.2 Limitations and 6.3 Future Work. There are 3 research questions proposed that can be answered by the currently available data:

RQ1: Can the Behaviour Analytics plugin cluster students properly according to their actions in the Learning Management System? This question can be answered by comparing the system generated clustering analysis to the manually altered result done by the teacher.

RQ2: Do the clustering results done by Behaviour Analytics and Teacher match the learning style found through Index of Learning Styles questionnaire? This question can be answered by comparing the system and teacher altered clustering results with the questionnaire results.

RQ3: Has the Behaviour Analytics plugin done a better job than Teacher on clustering students based on their actions in the Learning Management System? This question can be answered by comparing the difference between the results of the previous research question.

H1: The clustering result done by Behaviour Analytics compares against the final results adjusted by Teacher are similar.

H2.1: The clustering result done by Behaviour Analytics compares against the grouping based on responses of Learning Style Inventory questionnaire are similar.

H2.2: The clustering result adjusted by Teacher compares against the grouping based on responses of Learning Style Inventory questionnaire are similar.

H3: The clustering result done by Behaviour Analytics compares against the grouping based on responses of Learning Style Inventory questionnaire is closer than the clustering result adjusted by Teacher.

The literature review helps resolve these research questions by informing the design and implementation of the software tools used to explore the questions. From the literature review, novel and interactive visual learning analytics are needed, which the Behaviour Analytics plugin will provide. The sequence analysis articles informed the approach taken to obtain student clustering and the algorithm used, which is as a secondary analysis after conversion to a network graph and the k-means algorithm using log file extracted behaviours. The personalization articles informed the design of the software into the domain, student, and instructional model plugins.

The learning style prediction articles informed the choice of learning style model, which is the Felder-Silverman, as well as the choice of data, which is student behaviour patterns extracted from the log file. Also, comparing the prediction results with the questionnaire results is the standard practice to measure the accuracy of the prediction mechanism, so the current research will measure against questionnaire results as well. Finally, fuzzy clustering produced the best prediction results, so considering that a student could belong to more than a single learning style category may be necessary to achieve optimal results.

Chapter 3. Learning Object Relation Discovery based Behaviour Analytics

3.1 Learning Object Relation Discovery (LORD) plugin

This section will detail the LORD plugin that discovers the relationships between learning objects in a course. The purpose of the plugin, how it is used, as well as the inputs and outputs of the plugin will be discussed.

The LORD plugin is designed to assist the Behaviour Analytics plugin with graph node arrangement. The default graph in Behaviour Analytics groups the learning objects (i.e., nodes) based on their sections in Moodle and lets the physics engine of the D3 chart library determine the nodes final positions. It is then possible to manually rearrange the nodes if they don't seem grouped well. The manual arrangement is time consuming and based solely on human judgment. Ideally, a graph arrangement strategy should be automated and informed about the nature of the nodes so that a meaningful graph can be built, which is what LORD attempts to achieve.

The LORD plugin needs information about the nodes, or more precisely, the learning objects the nodes represent, to arrange the nodes in a meaningful way. Most learning objects contain text content that can be extracted and combined with natural language processing to discover the relationship between two learning objects. The strength of the relationship between the learning objects can then be used as a distance measure between nodes in the graph to group nodes that are similar repel nodes that are different.

The first step in the process is to extract the text content from the learning objects in the course. Unfortunately, there are a number of different types of learning object, and it is also possible for developers to create new learning objects. It is not feasible to deal with all the custom learning objects that are already available in the Moodle repository, let alone know what new types may be created in the future. Each learning object type has its own method of storing its text content, so for these custom learning objects, only a minimal amount of information is extracted. Regardless of type, each learning object has a name and an introduction, so this information is always available, even for custom learning objects.

There are also a number of learning object types that are built into Moodle and always available without the need to install them. These types will be available in all instances of Moodle, so is worth spending the time to extract as much text content as possible. For example, to extract the content of a 'Page' type learning object, the Document Object Model (DOM) is used to pull all the HTML heading and paragraph tags along with their encapsulated content. The same is done for the 'Book' type learning object and any 'URL' or 'File' type objects that return an HTML page.

The 'URL' and 'File' types are tricky to deal with because they could lead to anything. If the URL or File leads to text or HTML content, then the extraction is simple. However, URLs and Files could be PDFs, word processor documents, presentations, videos, images, or something else. When the URL or File is a document, AbiWord is used to extract its text content. AbiWord (https://www.abisource.com/) is word processing software that is free software, licensed under the GNU GPL, and readily available on GNU/Linux operating systems. AbiWord was chosen as a file converter because it runs from the command line and can convert PDF and word processor documents to plain text.

Other built-in learning objects include the 'Quiz' that contains questions, answers, and hints, all of which is stored in the database and just needs to be retrieved. Other learning objects that store content in the database include 'Wiki', 'Survey', 'Glossary', 'Lesson', 'Feedback', 'Forum', and 'Chat', among others. Each learning object type has its own database tables and scheme for storing its related data. With the built-in types it is possible to explore the database to determine how the data is stored and extract it. However, with unknown custom learning objects, the database structure is unknown, so it is not possible to extract more than the minimum name and introduction.

Once all the textual learning content is extracted from the learning objects, it needs to be separated in various ways. The first is to separate the names of the introductions from the rest of the content. The names of the learning objects are only compared with other names, introductions are only compared against other introductions, and finally, all other content is compared. The reason for this first layer of separation is because it is known that all learning objects will have names and introductions. Naturally, it makes sense to compare apples with apples and only compare names with names and introductions with introductions. Likewise, if the learning object has other data available for comparison and analysis, then these should be compared against each other as well. This means that, when making a comparison between learning objects, if the first has extra content and the second does not, then there is nothing to compare, other than the name and introduction.

The second level of separation happens at the paragraph and sentence level. When comparing two learning objects, each name is generally only a few words, no more than a sentence. The introductions and other content can range from multiple sentences to several paragraphs. To compare one sentence to another, each word in the sentence is compared against each other word. Likewise, paragraphs are compared sentence-by-sentence. So, each learning object must have its content broken down from bulk text into paragraphs, then from paragraph into sentences, so that each word in each sentence can be compared against each word in another sentence.

The LORD plugin extracts the text content from the learning objects and separates it into sentences. The actual comparison between sentences is handled by a third-party web service Word & Sentence Natural Language Processing Similarity Service (https://ws-nlp.vipresearch.ca/). It is the service that takes a key and target sentence as input and handles the word-by-word comparison, returning a similarity score between 0 and 1. A score of 1 means perfect similarity in that the sentences contain the same words. A score of 0 means no word in either sentence is similar to any word in the other sentence.

The comparison process provided by the similarity service can take some time to complete, many seconds or even minutes, with the length of the input sentences being the determining factor in how long the calculation takes. For this reason, the LORD plugin provides an option to restrict sentence length. There are also options to restrict how many paragraphs deep the comparison goes and how many sentences deep the comparison goes. For instance, it is possible to limit the sentence length to 10 words and only compare the first 2 sentences of the first 3 paragraphs. Alternatively, one could compare the first 3 sentences of the first 2 paragraphs of each learning object. These settings allow for some control over both the depth of comparison and the time required to complete the comparison of all learning objects. Options are also provided to weigh up the different comparisons. It is possible to weigh the name comparison more or less than the introductions or other content. The weight options allow for some control over which content causes the graph nodes to attract more.

The final level of separation is done at the word level. It was noted during testing that the similarity service did not utilize some words, generally referred to as stop words, such as 'a', 'the', 'for', and so on, but there were some others as well. A stop word dictionary was created which is initially populated with stop words from mySQL database. As each new word is encountered during the learning object relation discovery process, that word is checked for inclusion in the similarity service. If the word does not exist with the similarity service, then it is added to the stop word list. Otherwise, the word is added to the regular dictionary list. The LORD plugin also provides the option to add or remove stop words from the list. The stop word list is used to remove any unused words from the sentences before they are passed to the similarity service for processing performance efficiency.

As previously noted, the similarity calculation can take some time to complete, and subjects the relation discovery process to some limitations. The first is that the similarity service times out at two minutes, meaning that if the similarity calculation takes more than two minutes, it is aborted and the service returns no score at all. The option to restrict sentence length is useful when the calculations are taking too long. The second limitation is that the total time required to compare all learning objects with all other learning objects in a course grows exponentially with the number of learning objects in the course and with the amount of content to compare.

Running all the comparisons at the same time was deemed unfeasible. The PHP script calling the similarity service spends most of its time waiting and does not require many resources. However, even when waiting, PHP scripts have time limits as well and will not run until completion if all comparisons are made at once. Instead, the relation discovery process happens in pieces. First, the textual content is extracted and stored. Then, on each run, one learning object is compared with one other learning object. The comparison results are recorded, as are any errors from the service timing out.

The relation discovery process is implemented in Moodle as a scheduled task, meaning that it runs on a configurable schedule. Each time the task runs, be it once a day or once every 5 minutes, for each course the plugin is used, another step of the relation discovery process is done until completion. Even running every 5 minutes can result in waiting days or weeks for all the comparisons to be made. The LORD block shows the relation discovery progress including the number of comparisons that need to be made, the number that have been done, and the number of errors. After completion, there can still be errors due to the similarity service timing out. The errors are not reset automatically and need to be reset manually using the settings link in the LORD block. Usually, the sentence length will be reduced when resetting errors to lessen the chance of recurrence.

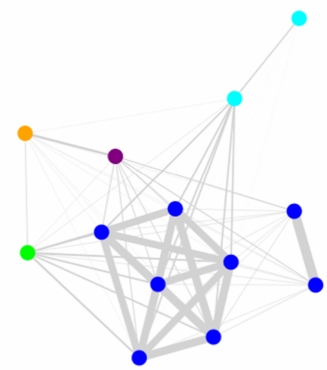
After the learning object relation discovery process has completed, a graph can be generated using the similarity scores between learning objects to inform the graph arrangement. While it is not necessary to wait until the relation discovery process is complete to generate a graph, the graph generation cannot be informed about the relations between objects until completion. The similarity score between each pair of learning objects is converted to a distance measure between graph nodes, where the greater the similarity, the smaller the distance and the lower the similarity the larger the distance. Converting the similarity measure to a distance value is necessary for its utilization in the graph generation process.

The D3 (Data Driven Documents) charting library (https://d3js.org/) provides the graphing functionality for generating and manipulating the network graph. The D3 library allows the graph links to have custom properties that can affect the node behaviour when the physics engine runs to produce the graph with its node arrangement. The links are given distance values so that when the physics engine tries to balance the various distances to produce a graph, it has the result of attracting nodes with a small distance and repelling nodes with a larger distance. The end effect is to produce a graph where the node learning objects are closer together if their learning content is similar and farther apart if their learning content is dissimilar.

The graphing interface in the LORD plugin has some options to control the final distance values, which affect the final graph. There are minimum and maximum node distances, which control how close the nodes are allowed to get to each other and how far apart they are allowed to be from each other. There is also a distance scale that is used to scale the similarity score to a distance. These three settings all affect how the graph looks when it is generated. However, if the graph is still not satisfactory, the nodes can be manually manipulated to have their positions changed. The plugin then offers the option to view the system generated graph or the manually altered graph. Figure 3 shows the completed graph for a Physics course along with the interface and information comparing two learning objects.

Physics 200: Introductory Physics I (Rev. C7)

Dashboard / Courses / PHYS200 (Rev. C7)



Use generated | Min distance = 40 | Max distance = 300 | Scaling factor = 800 | Regenerate graph | Allow changes?

Similarity: 0.03457889053333334

Module ID: 268
Name: Report an error or make a suggestion
Intro: You may send comments to the course coordinator regarding your learning experience in this course at any point.
P0: Feedback Comments to the course coordinator
P1: Feedback Your name and email (optional):

Module ID: 267
Name: Course Evaluation Survey
Intro: Complete and submit the Course Evaluation Questionnaire
P0: AU

Names 0.000377848900000000	report	an	error	make
course	0.000131	0.000051	0.001389	0.000020
evaluation	0.000002	0.000118	0.000000	0.000007
survey	0.000004	0.000057	0.000161	0.000001

Intros cost matrix 268 x 267 0.1033586982	0
0	0.1033586982
Optimal assignment:	0,0

Intros 268 S0 x 267 S0 0.103358698200000000	may	send	comments	to	the	course	coordinator	regarding	learning	experience
complete	0.000000	0.000002	0.000001	0.000097	0.000000	0.000000	0.000006	0.000038	-1.000000	0.000160
and	0.000000	0.028571	0.000005	0.000003	0.000000	0.000109	0.000556	0.033333	-1.000000	0.000135
submit	0.000008	0.000031	0.000001	1.000000	0.000000	0.000001	0.000118	0.000015	-1.000000	0.000321
the	0.000000	0.000000	0.000002	0.000013	0.000000	0.000000	0.000093	0.000001	-1.000000	0.000007
course	0.000000	0.000000	0.000001	0.000000	0.000000	0.000000	0.000000	0.000000	-1.000000	0.000001

FIGURE 3: COMPLETE LORD GRAPH WITH INTERFACE AND COMPARISON DATA

3.2 Behaviour Analytics plugin

This section describes the Behaviour Analytics plugin that uses the output of the LORD plugin. The purpose of the plugin, how it is used, its features, as well as its inputs and outputs are discussed. The Behaviour Analytics plugin is the main piece of outcome for this thesis. The plugin is designed to extract student sequential learning behaviour patterns in the form of log files from Moodle and use the patterns in clustering. Unlike some research that clusters log files directly, Behaviour Analytics clusters the log data only after some other manipulation. Before the clustering can begin, a network graph is made where the nodes in the graph represent the learning objects in a course. The student log data is mapped onto the network graph and then compressed into a centroid point with x and y coordinates that make clustering simple. The plugin offers a novel interactive visualization for clustering analysis.

The original graph produced by Behaviour Analytics has the nodes corresponding to learning objects in the course, where each node is coloured based upon its learning object type. The links are used to connect the nodes to their respective sections and hold the graph together but serves no other purpose. The original graph groups nodes by section, which may or may not be an appropriate configuration. Therefore, it is also possible to manually manipulate the nodes and produce a custom graph configuration. For the courses that have lots of learning objects needing to be positioned, the process is time consuming and tedious. An automatic configuration would be preferred.

The LORD plugin, previously described, was built to provide Behaviour Analytics with an automatically configured graph. The LORD graph is generated based on similarity between learning objects, where similar nodes are grouped together and dissimilar nodes are pushed apart. Behaviour Analytics provides options for integration with the LORD plugin, when LORD is installed. Either the system generated graph or the manually manipulated graph from LORD can be used with Behaviour Analytics. As shown in Figure 4, The LORD plugin provides a meaningfully and automatically configured graph to replace the original. Other informed graph generation procedures may also be employed in a similar way.

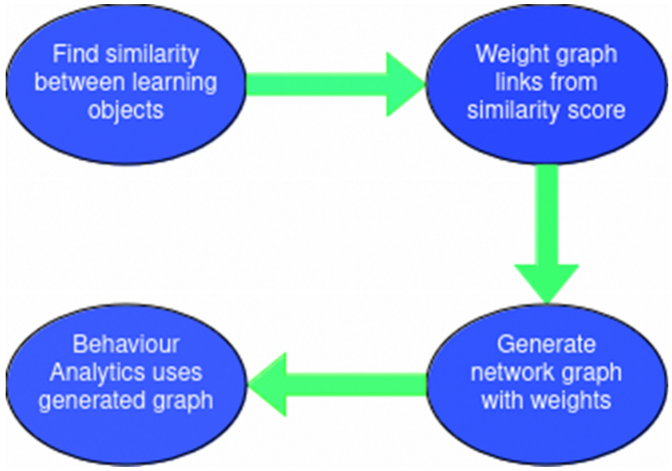


FIGURE 4: LORD INTEGRATION WORKFLOW

Whichever graph teachers decide to use, the plugin provides an interface for manually arranging the nodes. The interface consists of a hierarchical menu with check boxes to add or remove individual nodes or entire sections. There is also a weight slider that controls the attractive force in the links. Whenever a node is being dragged, the physics engine is running and the nodes may adjust their positions according to the current physics parameters. A higher number on the slider means the nodes are more attracted to each other, while a negative value means they repel each other. A value of 0 indicates the other nodes will remain stationary while one is being dragged. Figure 5 shows the interface for manually configuring the learning object graph. The figure has a list of users on the left side that is available to those with the researcher role, but teachers will only see their own graph.

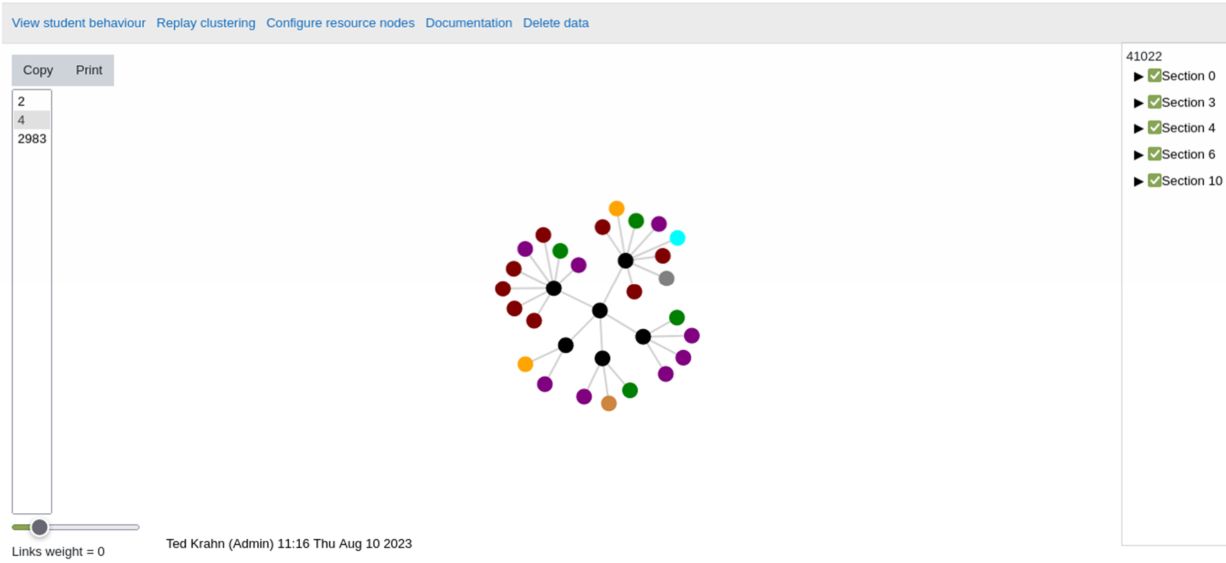


FIGURE 5: BA GRAPH CONFIGURATION INTERFACE

The second interface that Behaviour Analytics provides is 'View student behaviour' that is accessed via a link in the block (see Figure 6 below). The viewing page shows a list of groups being used in the course and a list of students along the side, a double handled slider along the bottom, and the graph in the centre. From this interface, it is possible to select one or multiple students from the list, which will show the students' mapped log file data on the graph. Each student has their own colour of links and multiple students' behaviour can be viewed at the same time.

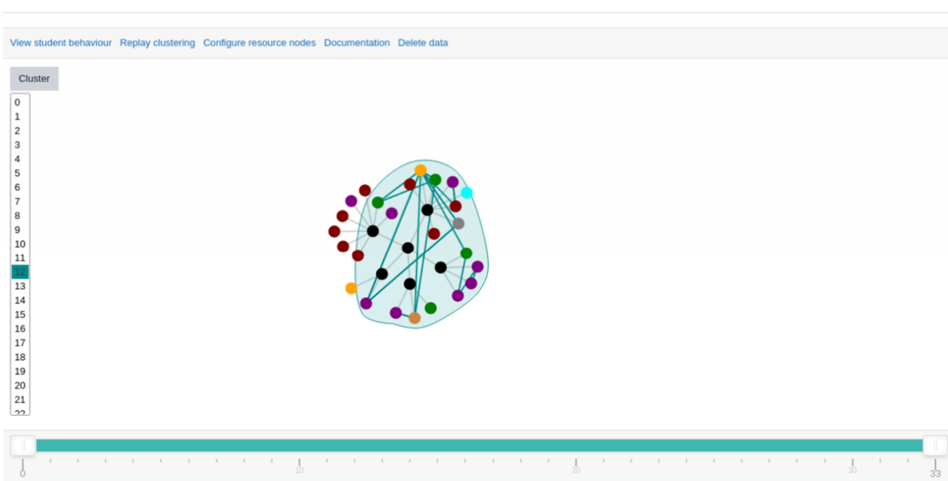


FIGURE 6: BA VIEWING INTERFACE WITH STUDENT DATA SELECTED

The slider along the bottom denotes the number of links available, where one handle is at 0 and the other at the maximum. All students have their data shown according to the values on the slider. Each student's data starts at their first link and progresses until their last link. Moving the 0 handle shows the student behaviour starting from link 1 or 2 or 3 rather than 0. Likewise, moving the maximum valued slider handle results in the viewing of student data from link 0 (or whatever the value happens to be) to near the end, less the second slider handle value. All students will have data at the 0 end, but not all students will have data at the maximum end. The slider provides a way to view a time slice of the data.

The viewing interface also contains a 'Cluster' button at top left-hand corner. With students selected and a time slice set, clicking the button will show the clustering interface. The clustering interface (see Figure 7) consists of a log panel that displays running text of the clustering analysis, a set of control buttons to play and pause the clustering, and a slider to control the stages of the interface.

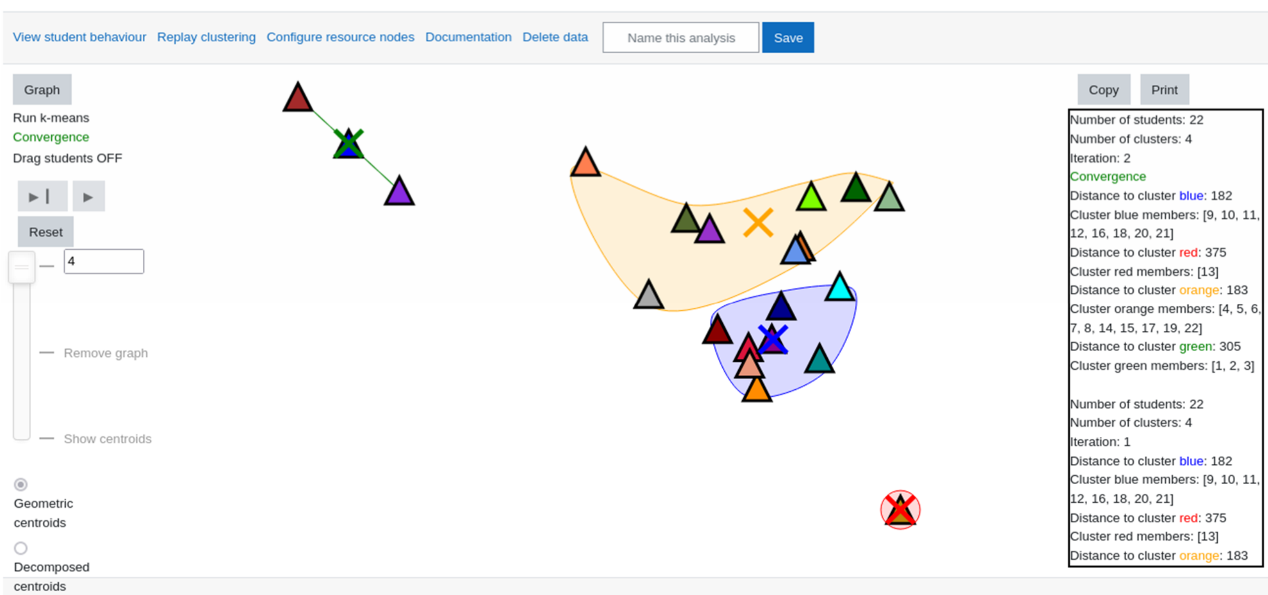


FIGURE 7: BA CLUSTERING INTERFACE AT ANALYSIS CONVERGENCE

The first stage of the interface shows the graph and student links with each student's data represented as a triangle, which is the student's centroid derived from the coordinates of the nodes they visited as per the log data. The second clustering stage removes the graph and scales the student centroids to fit the available screen space. The third stage allows the choice of number of clusters and allows the clustering process to begin. The clustering process is controlled by a play button and a play/pause button. Clicking the play button will start the clustering process and continue until convergence. Clicking the play/pause button steps the clustering process forward a single iteration at a time. When playing, the clustering can be paused and resumed. There is also a reset button that reverts the interface back to the first stage. Figure 7 shows the clustering interface at the completion of a clustering run where all 4 clusters have converged.

Additional features available during clustering include the ability to attach a comment to a student or cluster centroid, ability to drag and drop student centroids to include them in a particular cluster, and hovering over a cluster centroid will display a graph with the common link among those members. Attaching comments to student and cluster centroids allows observations made by the teachers doing the clustering to be passed on to the researchers evaluating the plugin performance. Dragging and dropping students into different clusters can help the clustering achieve visually satisfying results, as the clustering algorithm (k-means) is not concerned with appearance, only the calculations of its algorithm. The clustering analysis, at this stage, will not allow the manually placed students to override the mathematical calculations of the clustering algorithm and will only converge when its calculations tell it to.

Finally, the common links among members of a cluster are part of the output of Behaviour analytics. These common links will be used to predict the learning style of the student from their behaviour and will inform the Personalised Study Guide about how to adapt the course learning objects. Hovering over the cluster centroid allows these common links to be seen. It can happen that it appears there are no common links, when there are, as some links are linked to themselves, which happens when the student clicks the same learning object twice in a row. If such self-linked links are the only common ones, then there may be common links even in the appearance of their absence.

The last interface that Behaviour Analytics provides is the clustering replay, which consists of the same log panel as the initial clustering interface, play, pause, and stop controls, a button to delete the selected dataset, and a select menu of the various clustering analyses that have been done through the initial clustering. The menu will only list LORD graph-based clustering analyses when the plugin is configured to use LORD and will display other graph analyses otherwise. Figure 8 shows the replay interface with the same clustering as in Figure 7, but with some manual clustering done and the common links graph of the blue cluster.

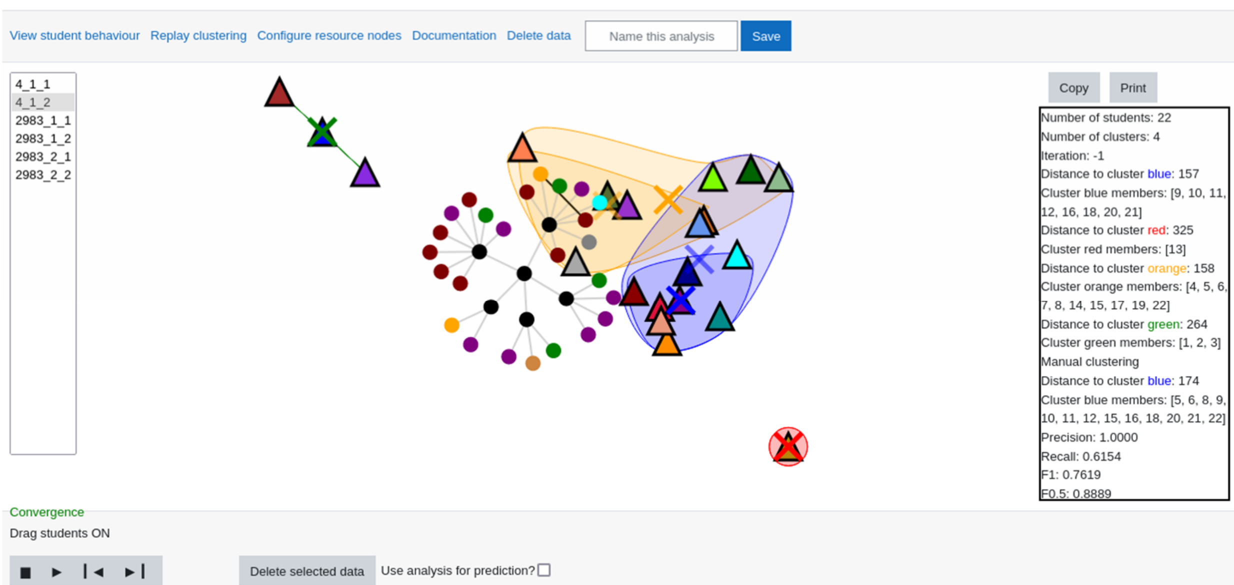


FIGURE 8: REPLAY INTERFACE WITH MANUAL CLUSTERING AND COMMON LINKS (BLUE)

Selecting an analysis will display the graph with student links and student centroids. Each analysis can be replayed from iteration 1 until initial convergence, and if extra data is present for that analysis, beyond initial convergence. During replay, it is also possible to click student and cluster centroids to attach comments, same as in the initial clustering, with the difference being that there may already be comments for that student or cluster, so these comments will show and can be altered during the replay stage.

When a clustering analysis is done initially through the viewing interface, the same analysis may be updated as new student data is available. An analysis will get updated only if the time slider in the view interface before initial clustering had its handles at 0 and maximum. Running a clustering analysis on a different time slice is possible and will show up in the replay menu but will not be updated as new student data becomes available. The reason for not updating different time slices is that it is assumed that someone viewing a time slice will want only that data used in the clustering analysis, so there is no updating with new data. However, when someone selects 0 to maximum, this indicates that they want all available data used in the analysis, so it is updated as new data are available.

The time slider allows teachers to see different slices of time of their students behaviour. This feature allows the teacher to ignore early behaviour where students may be exploring the course before working through the material. Alternately, the teacher could view just that initial behaviour to see if the students are exploring all the learning objects before working through the material.

The replay interface allows the updated clustering analysis to be seen. When the clustering is first running, the iterations start at 1 and increase sequentially until convergence. When the initial clustering reaches convergence, the iterations become negative, starting at -1 and decreasing sequentially. Using negative numbers allows for a distinction between the initial clustering iterations and clustering iterations done post-convergence with new data. When replaying an analysis with updated data, the student centroids will often move around screen from iteration to iteration, indicating that the student has new data, which is causing a change in their behaviour graph, which alters their centroid coordinates.

When the student centroid moves, it may become part of a new cluster and take on a different set of common links. The updating of analyses allows for students to change clusters and common links, which in turn will cause their learning style to be predicted differently and will affect how their Personalised Study Guide is arranged. The updating process allows the student learning style prediction to become dynamic rather than remaining static, as with a questionnaire, and can adapt to changes in observed student behaviour.

The replay interface offers one other important feature, which is the ability to manually adjust the clustering results. Manually changing the clustering memberships by dragging and dropping student centroids was possible in the initial clustering, but there the k-means calculations would override the manual clustering to reach convergence. However, during replay, once the replay is past initial convergence, the manual clustering will remain in place. In fact, if one changes the clustering membership during replay, there will become two sets of clusters, one generated by the plugin's k-means implementation and the other generated manually by the user.

The manual clustering during replay is essential to the performance evaluation of the plugin. Teachers using the plugin to cluster their students will observe the system clustering results and make any adjustments to the membership they deem appropriate. The manually generated clusters are taken as correct and compared with the plugin generated clusters to determine the precision, recall, and F-measures. These measures are noted in the log panel when there is manual clustering data to compare against. For Behaviour Analytics to be considered useful, it will need to produce system generated clustering that is similar or identical to the clustering done by domain experts.

3.3 Performance Analysis

This section discusses the performance of both LORD and Behaviour Analytics. The LORD plugin uses natural language processing (NLP) to discover the relationships between learning objects, which is a computationally expensive task and will be described. The Behaviour Analytics plugin uses clustering that also comes with a computational price, which will be described.

The LORD plugin relies on Natural Language Processing (NLP) to determine the similarity between learning objects. However, the LORD plugin relies on a third-party service for the computationally expensive sentence similarity measure and spends more time waiting for the service to return a result than doing any processing itself. LORD does extract the learning content from the learning objects, which may be a simple matter of pulling data from a database or may be slightly more complex such as parsing text from a PDF or a Web page. The content extraction does not represent any kind of a bottleneck for the program.

The learning object comparison stage takes the most time with LORD, but it is mostly idle and spends time on waiting. When comparing learning content, there may be many paragraphs that need to be compared, each paragraph containing many sentences. The comparison process involves checking every sentence in every paragraph against every other sentence in the other paragraph, up to the maximum number of sentences and paragraphs to compare. The comparison for each set of content then has a $P^2 * S^2$ complexity, where P is the number of paragraphs to compare, and S is the number of sentences to compare. When the paragraph and sentence setting values are high, the comparison will take longer.

The LORD plugin does not do the word-by-word comparison, as this is done by the web service. However, the service has a W1 * W2 complexity, where W1 is the number of words in the first sentence and W2 is the number of words in the second sentence. The word-by-word comparison is the most time-consuming part in the comparison process, so the setting that has the most direct influence on time spent is the sentence length. The total complexity of the comparisons is $P^2 * S^2 * W1 * W2$, where W1 and W2 will often vary between sentences.

Setting the sentence length to a very low value like 1 or 2 has the benefit that comparisons happen much faster, but at the cost of accuracy. Such low sentence length will not produce quality comparisons and realistic similarity values. The default value for sentence length is 32, which worked well in testing, and it will ensure that most sentences will get compared in their entirety. If the web service times out and returns nothing for a comparison, the LORD plugin records this as a comparison error. After all the comparisons have been made, the sentence length can be reduced and the errors reset, so that those comparisons with errors can be redone with the lower sentence length, which will increase the likelihood that the comparison completes successfully and does not cause error again. The Behaviour Analytics plugin performs everything on its own, so spends no time on waiting for third-party services. However, because it does everything, slow operation cannot blame elsewhere. To minimize server processing, the plugin attempts to offload much of the work to the client machine by using JavaScript, which is used by the D3 library (used for all the graphing) and all the interface components. The server also processing data when necessary but does not handle the interface.

When Behaviour Analytics is first used, there is no graph generated, unless LORD is used. The original graph is made entirely by the client machine and has no node or link data until after it has been rendered. The graphing process consists of creating the nodes and links, then letting the D3 physics engine run a few iterations where it tries to balance the physics forces at work. Small graphs of a few nodes pose no issue, but if a course had an excessive number of learning objects, the number of calculations necessary each iteration could start to slow down the graph generation process. During testing, a course with about 120 learning objects was used and the graph generated without issues, so there would need to be thousands of learning objects in a course before noticeable problems occurred. But, that many nodes in a graph would also consume the entirety of the available screen space, causing another problem.

When viewing student’s behaviour, clicking a student in the menu displays that student's behaviour graph. Each student also has a hull drawn around their graph, which is a shaded area encompassing their links. As with the graph generating process, a few students with a few data points do not cause any issues, but when many students are selected that have a lot of log data to display, the calculations to generate all the links and hulls creates a degraded visual effect. During testing, student data was imported multiple times to create a small selection of students that each had thousands of log entries. It was noticed that the visual effects in the graph start to become delayed and do not show smoothly as they do otherwise. The hull calculations consider all the node coordinates in each student graph, so when many thousands of data points are needing to be processed, a bottleneck is produced.

The Behaviour Analytics plugin also implements the k-means clustering algorithm in both JavaScript and PHP. The clustering phase uses an implementation that can be stepped through one iteration at a time, played from start to finish, or played then paused. The use of stepping through the clustering is that it becomes interactive. Between iterations, it is possible to add comments to centroids or drag and drop students to change their cluster membership. Such function and user experience would not be possible when the clustering algorithm is run in the background and only shows the results. Stepping through the iterations does not affect the performance of the algorithm, as it is still k-means. At each iteration, each student has their centroid's distance from each cluster centroid calculated so they can be assigned membership to the nearest cluster. The complexity of the k-means algorithm is I * C * S, where I is the number of iterations, C is the number of clusters, and S is the number of students. The PHP implementation of k-means provides no visual as it runs in the background as a Moodle scheduled task. The implementation differs because it does not step through the iterations and deals with a larger dataset, but the required calculations and run-time complexity is the same.

K-means is not the most efficient clustering algorithm, but it is simple, popular, and easy to implement, which is why it was chosen. The run-time of k-means indicates that it is not a big data solution, and indeed much more efficient clustering algorithms exist for big data sets. Behaviour analytics also has other features, which have been noted, that indicate it is not a big data solution. The plugin was designed for use in an educational setting where it deals with a single course worth of students and student data at a time. The computational complexity of its visual features and use of a brute force clustering algorithm means the functionality used in the plugin may have limited application outside its current context.

Chapter 4. Personalized Study Guide

4.1 Learning Style and Learning Object

The Personalized Study Guide (PSG) Moodle plugin is explained in details in this section. Some background information will first be provided to understand what the PSG plugin is attempting to do. The plugin uses the concepts of learning styles and learning objects to personalize the learning environment. Learning styles and learning objects will be described to provide relevant background information before the plugin itself is examined.

This research aims to use FSLSM learning styles to adapt a Moodle course Web page layout to suit the style of the learner. The course page contains a few resources of various types such as Forum, Assignment, and Quiz which are the learning objects for the course.

There is no established definition of learning object and various standards exist for learning objects that can be adhered to for reusability/interoperability. One definition of a learning object is provided by McGreal (2009):

> "...any reusable digital resource that is encapsulated in a lesson or assemblage of lessons grouped in units, modules, courses, and even programmes. A lesson can be defined as a piece of instruction, normally including a learning purpose or purposes."

With this definition, McGreal attempts to synthesize other definitions and provide one that is neither too limited in scope nor too broad. Moodle courses with their sections and modules are digital, reusable, and have a learning purpose, so fall under the aforementioned definition. This research considers each course module or activity to be a learning object to be rearranged, but the PSG plugin can also rearrange sections as well. When section order is adapted, each section will retain the activities that were originally in it. Figure 9 shows part of a Moodle course page with a selection of learning objects. In order of their appearance top to bottom, the object types are file, chat, URL, page, choice, and forum. The order of the learning objects is adapted to suit the learning style of the student.

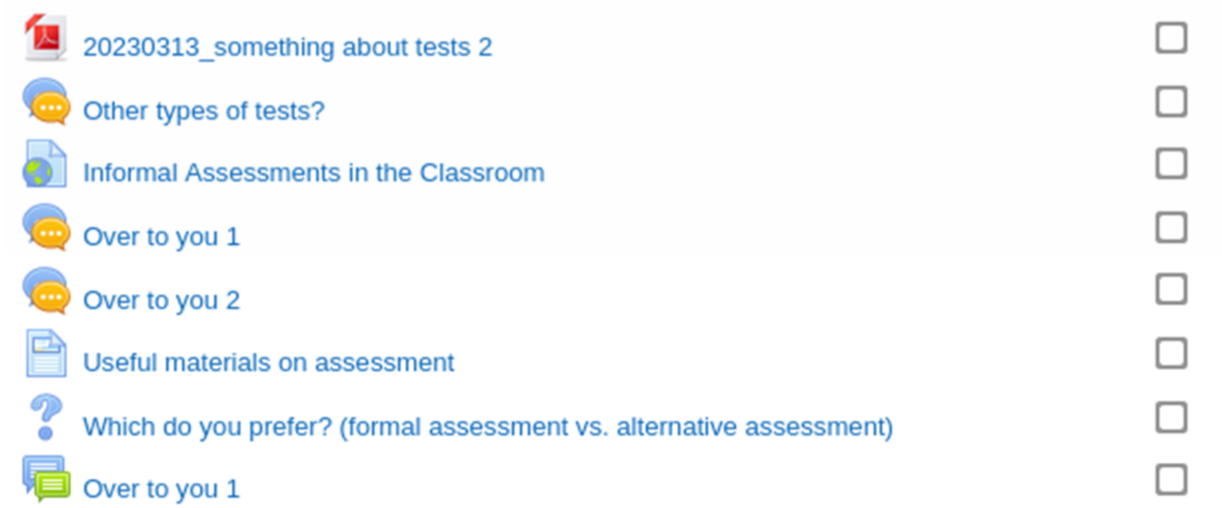


FIGURE 9: SECTION OF A MOODLE COURSE WITH ITS LEARNING OBJECTS

4.2 Architecture and Workflow

The Personalized Study Guide (PSG) Moodle plugin is a course format type plugin designed to replace the format of a course using the Topics format. If it is used in a course made with a format other than the Topics format, then the PSG may not work correctly. The Topics format contains a few sections that each contain one or more learning objects. The PSG can reorganize both the sections and the learning objects within those sections but does require that the Behaviour Analytics (BA) plugin is first installed and configured. The BA plugin determines the learning object and/or section personalization by providing ILS scores through its survey management system or by analyzing the learning objects among common links in a BA clustering analysis. If BA does not provide any information to PSG, then there is no personalization. Figure 10 shows the workflow of the LORD, BA, and PSG plugins.

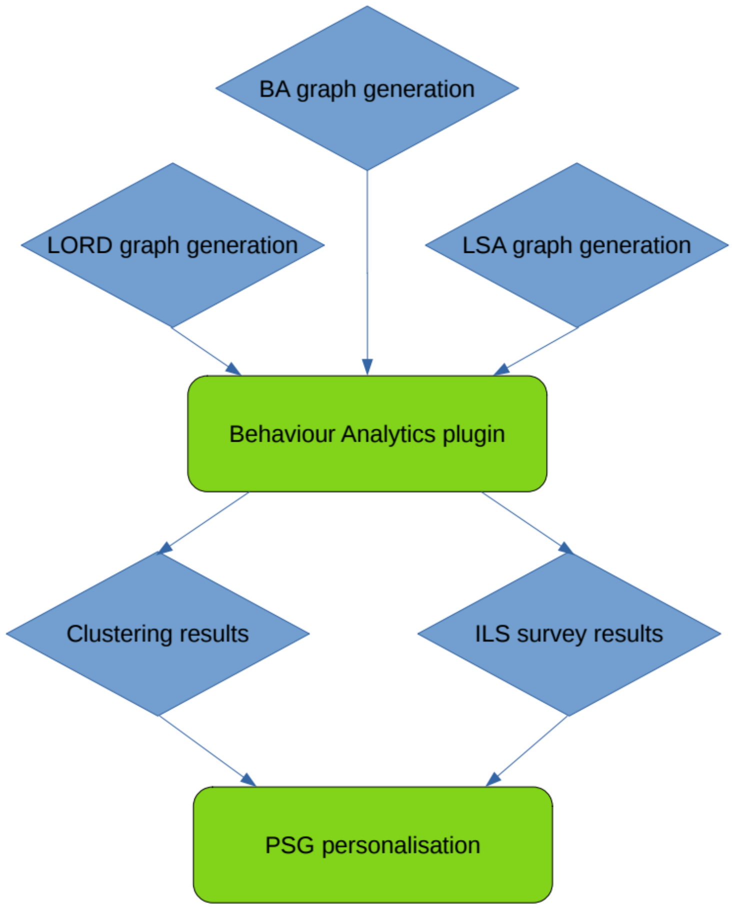


FIGURE 10: PSG WORKFLOW

If a student is using the PSG and has filled out the ILS, then the course page can be personalized for that student's learning style. Each learning object has a learning style derived from the type of learning object it is. For example, an image would be visual rather than verbal, a book would be verbal rather than visual, and a forum or chat would be active rather than reflective. Therefore, each learning object and each student has a learning style combination comprised of the dominant FSLSM dimensions. The student's combination and the object's combination are compared against each other looking for common dominant dimensions producing a score. A learning object may score 0 if it has no common learning style dimensions with the student and can score 4 if all four dimensions are aligned. Each learning object on the course page is then ranked according to its alignment with the student's learning style and shown in ranked order. Sections are scored and ranked based on the total score from all the learning objects in that section.

If BA is providing common link data rather than ILS results, then there must be a clustering analysis done and selected for prediction within that course. The learning objects comprising the common links of the cluster in which the student resides are used to determine that student's learning style. Each of the objects in the common link group has its learning style score multiplied by the number of times that object was used by the students in that cluster. If a learning object was used 3 times by all members of the cluster, then that learning object is 3 times 'heavier' in the final learning style calculation for the cluster. It is the learning style combination of the common links within a student's cluster that are given to PSG to determine the personalization.

The use of common links allows for the learning style results of the cluster to change over time. A clustering analysis that has been done before the students are finished with the course will be updated as the students interact with the course learning objects. The BA plugin uses a scheduled task to look for new object usage data and updates any clustering analysis to reflect the new data. Students may change clusters and common links are recalculated during such updates.

The personalization of the course material aims to answer the second research question. By changing a regular course into a Personalized Study Guide it is hoped that the students using the PSG will benefit from the personalization that puts their preferred learning objects to the top of the section and the section with the most preferred learning objects at the top of the course page.

4.3 The Personalized Study Guide Plugin

The PSG plugin is activated by altering the course settings to change the course format. When the PSG is active for a course, then the BA block will have a button at the top to turn personalization on and off, which gives the student the ability to use the course as the designer made it or a personalized version of the same learning material. In the BA dashboard it is possible for a teacher to see when and how often a student changes the personalization on and off to better understand how the students are using the personalization.

As has been mentioned previously, the PSG can personalize both the learning objects within a section and can also the sections themselves. It is also possible to use either ILS results to determine learning style or common links from a cluster. These options are controlled through the plugin settings, which are only available to administrators due to the nature of the plugin. Figure 11 shows the settings page for PSG. Figures 12 and 13 show a course using the PSG with personalization both off and on. In Figures 12 and 13 both the learning objects in each section and the sections are being personalized.

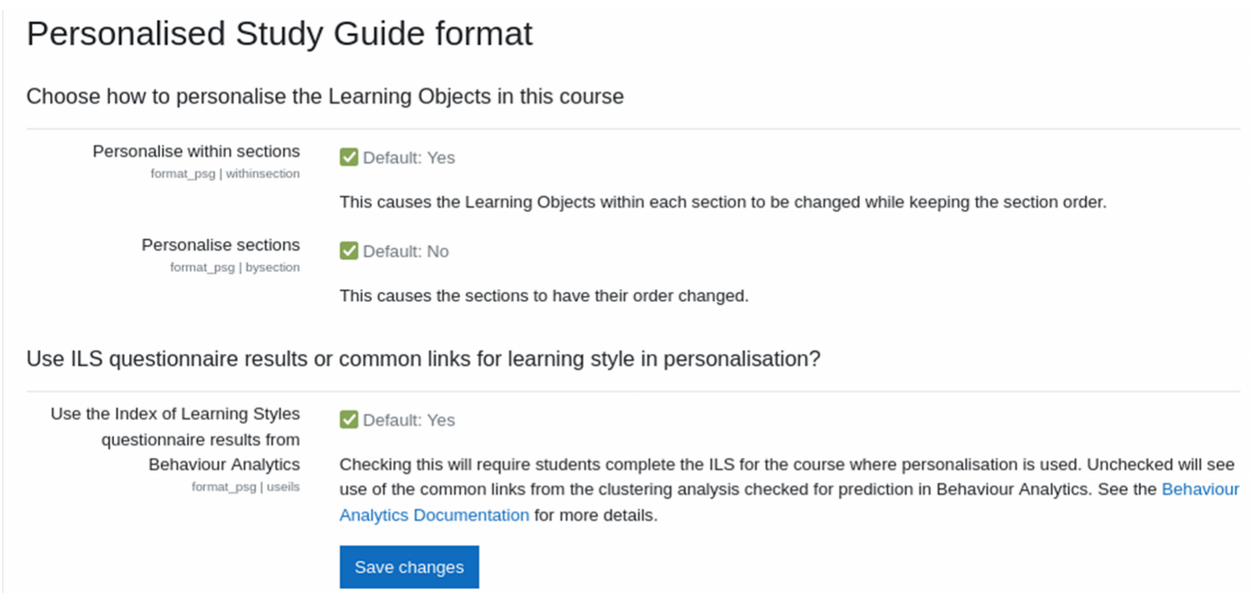


FIGURE 11: PSG SETTINGS PAGE

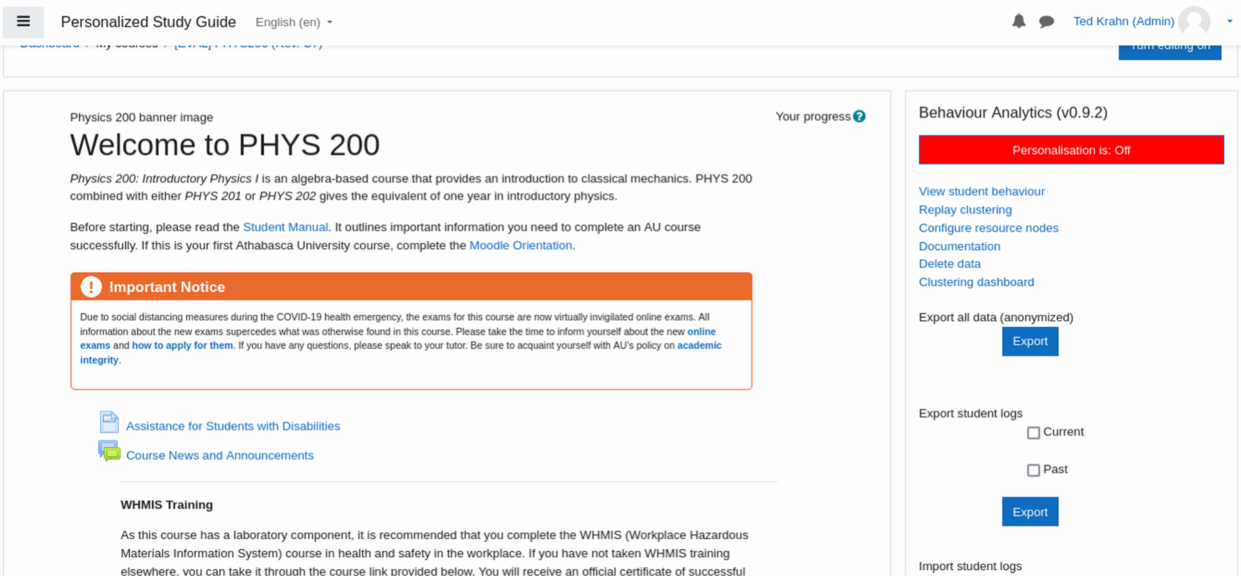


FIGURE 12: PSG COURSE WITHOUT PERSONALIZATION

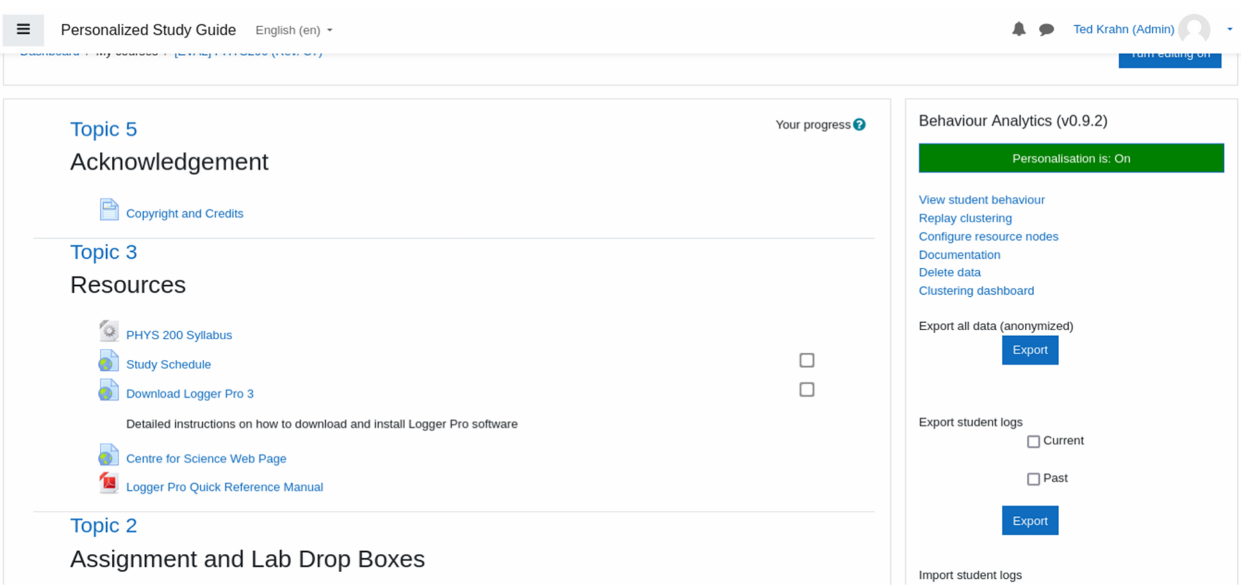


FIGURE 13: PSG COURSE WITH PERSONALIZATION

Chapter 5. Evaluation

5.1 Evaluation Plan

There were two research questions proposed in Section 2.5, which were about clustering to detect learning styles and personalized learning using learning styles. However, only the data collected can be analyzed and there is no personalized learning data available at the time of this writing. The data that is available drives this research to consider an alternative set of questions elaborating on the first original question. The data available are a set of ILS results and a system generated manually altered clustering analysis for a small class of students.

The first original research question asked: Can the log file clustering provided by the Behaviour Analytics plugin detect student learning styles? Since the data collected includes ILS results with a clustering analysis, the question can be explored by deriving three more specific research questions from the original:

RQ1: Can the Behaviour Analytics plugin cluster students properly according to their actions in the Learning Management System? This question can be answered by comparing the system generated clustering analysis to the manually altered result done by the teacher.

H1: The clustering result done by Behaviour Analytics compares against the final results adjusted by Teacher are similar.

RQ2: Do the clustering results done by Behaviour Analytics and Teacher match the learning style found through Index of Learning Styles questionnaire? This question can be answered by comparing the system and teacher altered clustering results with the questionnaire results.

H2.1: The clustering result done by Behaviour Analytics compares against the grouping based on responses of Learning Style Inventory questionnaire are similar.

H2.2: The clustering result adjusted by Teacher compares against the grouping based on responses of Learning Style Inventory questionnaire are similar.

RQ3: Has the Behaviour Analytics plugin done a better job than Teacher on clustering students based on their actions in the Learning Management System? This question can be answered by comparing the difference between the results of the previous research question.

H3: The clustering result done by Behaviour Analytics compares against the grouping based on responses of Learning Style Inventory questionnaire is closer than the clustering result adjusted by Teacher.

The second original research question asked: Can providing students a personalized study guide based on Behaviour Analytics improve the students' learning experience? There is no data available at this time to explore the question and the lack of data is both a limitation and a future work (see Sections 6.2 and 6.3).

5.2 Data Collected

5.2.1 Responses to Index of Learning Styles Questionnaire

The data was collected from a course entitled "Intermediate Listening & Speaking Practice (II)" from the National Changhua University of Education in 2023. There were 22 students in a class of 27 who opted to fill out the ILS. Of the 16 possible learning style combinations available with the FSLSM, only 5 different learning styles were present. For the active/sensing/global/visual, active/intuitive/global/visual, and reflective/sensing/sequential/verbal learning style categories there was only 1 student in each category. There were 8 students with the reflective/sensing/sequential/visual learning style and the other 11 students had the active/sensing/sequential/visual learning style. The student response data is summarized in Table 15.

TABLE 15: STUDENTS BY LEARNING STYLE CATEGORY

Learning Style Category	Student IDs
active/sensing/global/visual	0
active/intuitive/global/visual	9
reflective/sensing/sequential/verbal	20
reflective/sensing/sequential/visual	2, 5, 6, 10, 11, 12, 17, 18
active/sensing/sequential/visual	1, 3, 4, 7, 8, 13, 14, 15, 16, 19, 21

The students all share the sensing and verbal attributes, with a single exception in each case, and all but 2 students are sequential learners. The similarities in learning style classification may be due to the type of course used where similar students have chosen the same course or program to take. More information about the students is necessary to determine why these similarities exist among the class.

The BA dashboard shows radar charts when viewing the ILS results for a course. Figure 14 shows the radar chart for the average learning style of the 22 students. It can be seen from the chart that this sample of students has preferences for the sensing, visual, and sequential dimensions.

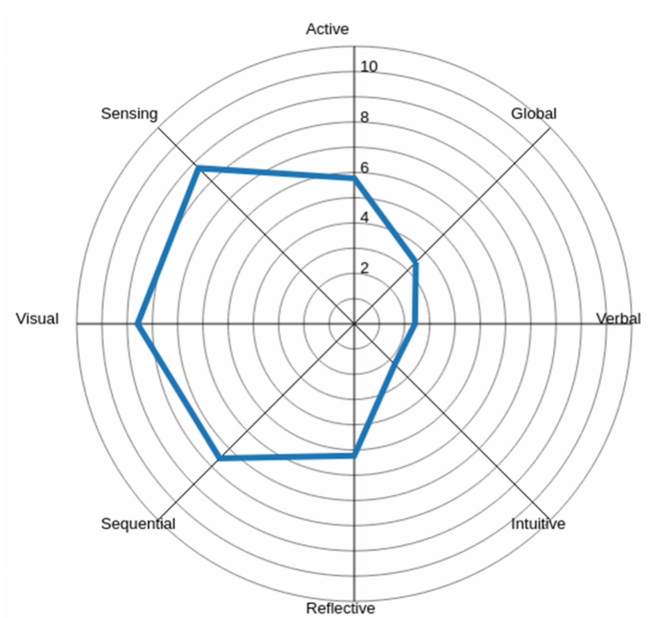
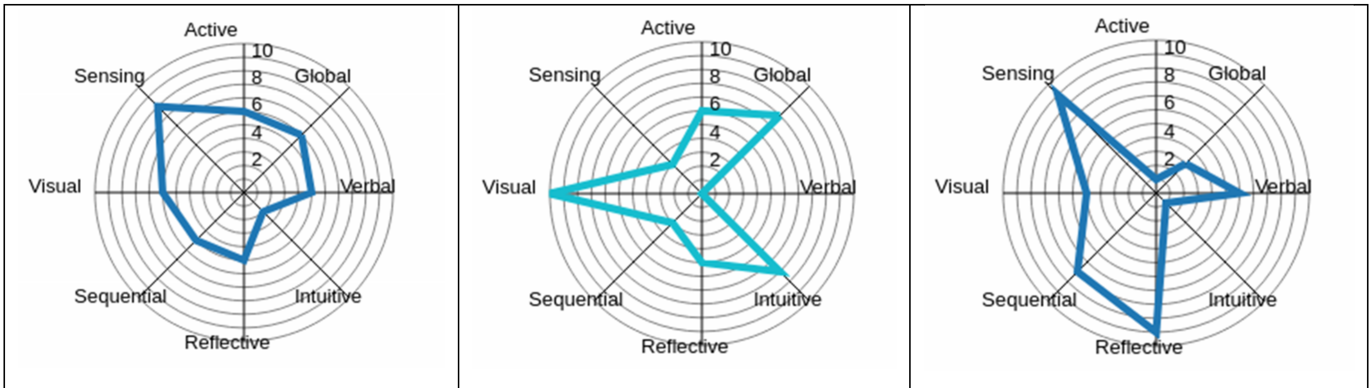


FIGURE 14: SPIDER CHART OF AVERAGE LEARNING STYLE

The dashboard also shows radar charts for each student. The charts for each of the 5 learning style categories are shown in Figures 15 to 17. The categories are presented in the same order as before with the 3 single student categories first (see Figure 15), then the 2 larger groups (see Figures 16 and 17). It should be noted that while a student may belong to a certain group, they may have strong, medium, or weak preferences for any given dimension.



Active/Sensing/Global/Visual	*Active/Intuitive/Global/Visual*	*Reflective/Sensing/Sequential/Verbal*

FIGURE 15: THREE SINGLE STUDENT LEARNING STYLE GROUPS

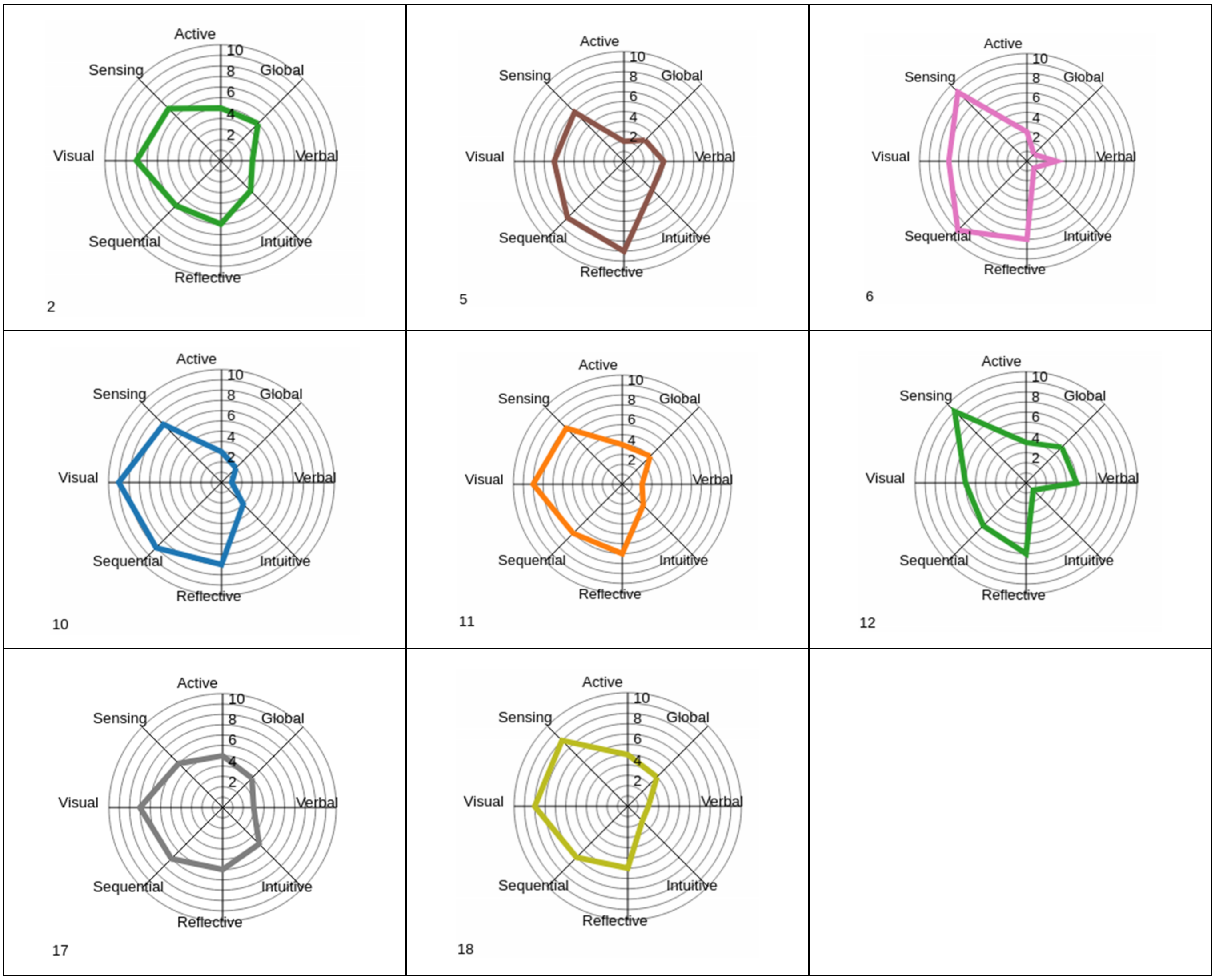


FIGURE 16: REFLECTIVE/SENSING/SEQUENTIAL/VISUAL STUDENTS

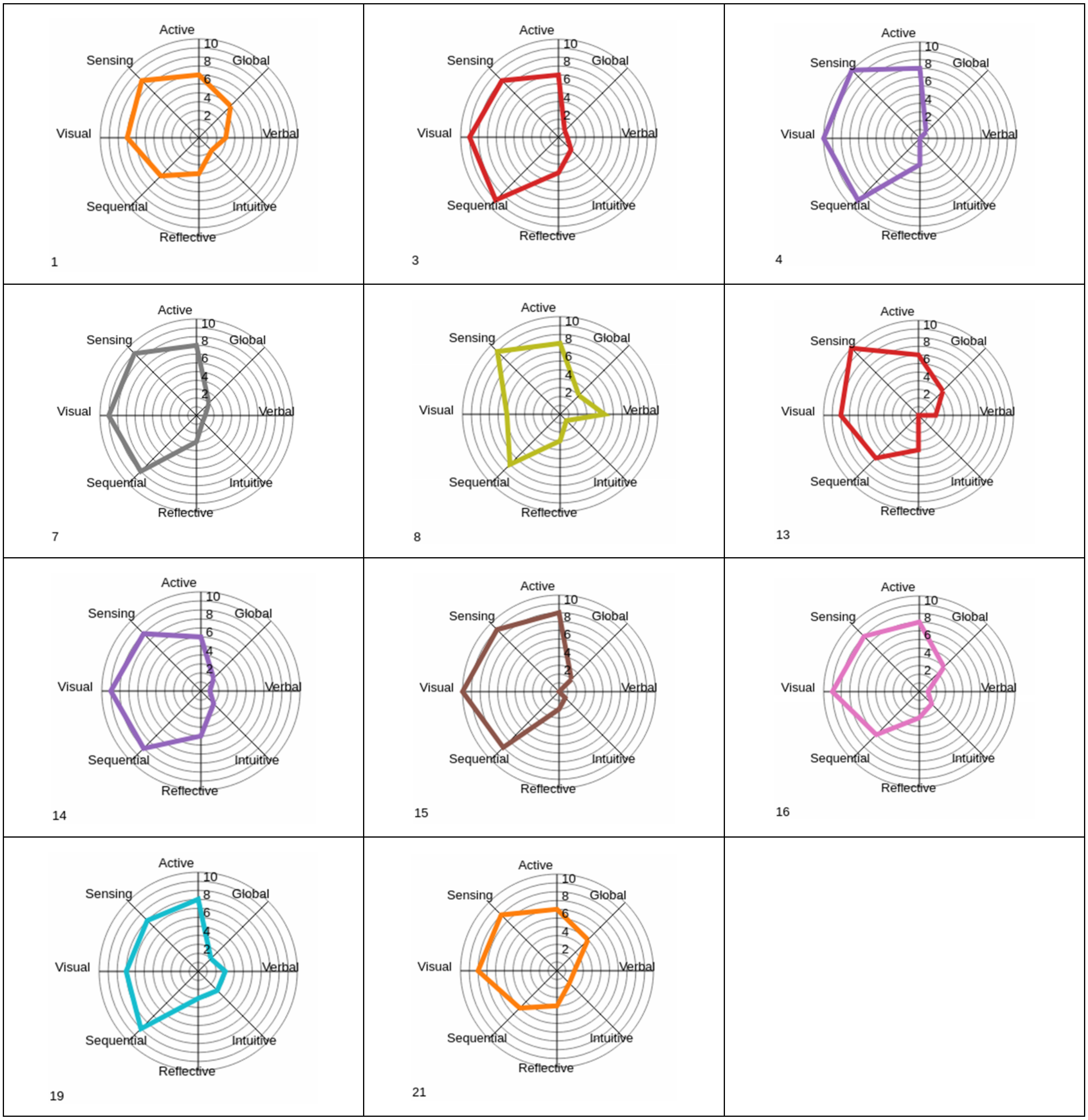


FIGURE 17: ACTIVE/SENSING/SEQUENTIAL/VISUAL STUDENTS

5.2.2 Learning Object Graph and Clustering

With a clustering analysis done, it is possible to examine the Learning Object Graph (LOG) for each of the groups to see if there are any obvious patterns. By only selecting the students in a given cluster in the BA viewing interface, the various LOGs can be viewed at the same time and compared. There are 5 clusters in the analysis, but cluster #1 is admin click data that can be ignored. Of the other 4, a cluster stands out from the remaining 3. There are 3 students who interacted with the course learning objects very little and none of these 3 took the ILS. It is assumed that they failed or dropped out from the course. Figure 18 shows the LOGs of these low interaction students. One student clicked two learning objects while the other 2 students clicked one each.

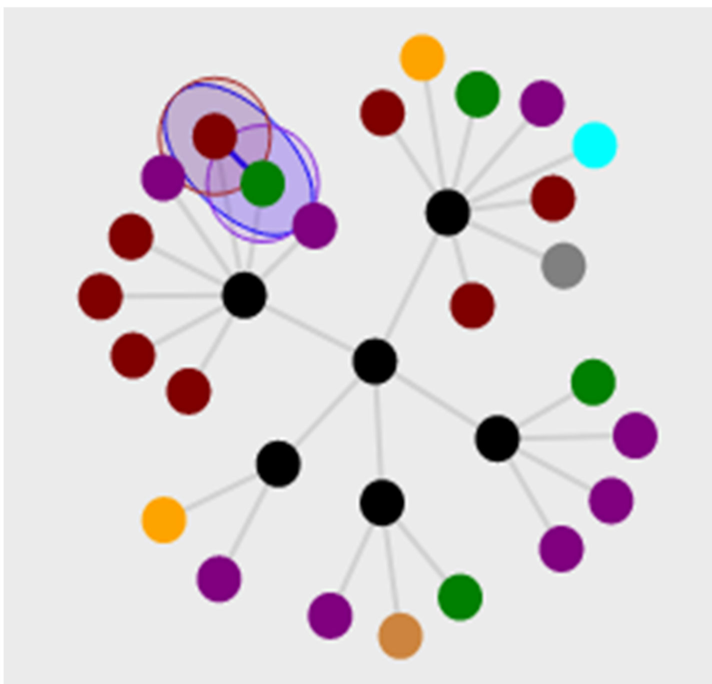


FIGURE 18: LOW INTERACTION CLUSTER LOGS

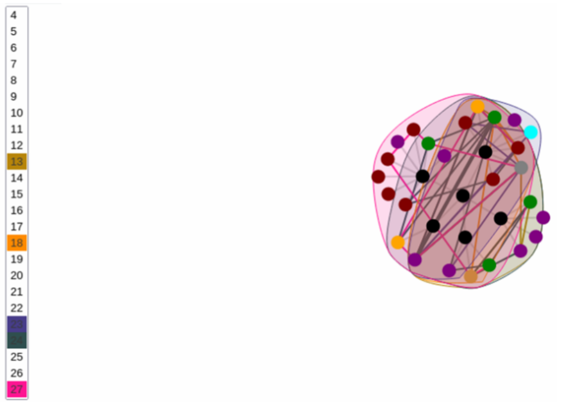
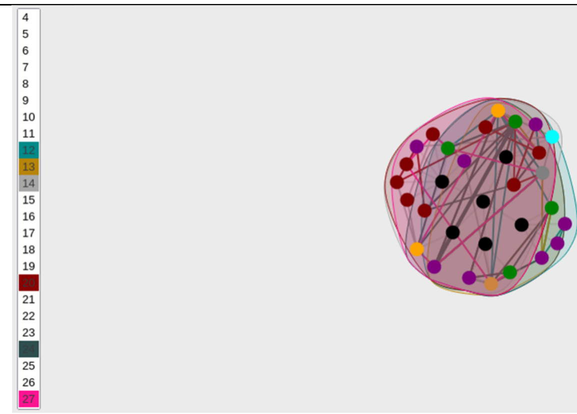
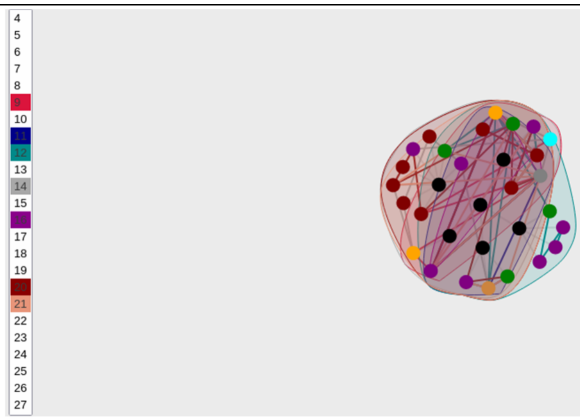
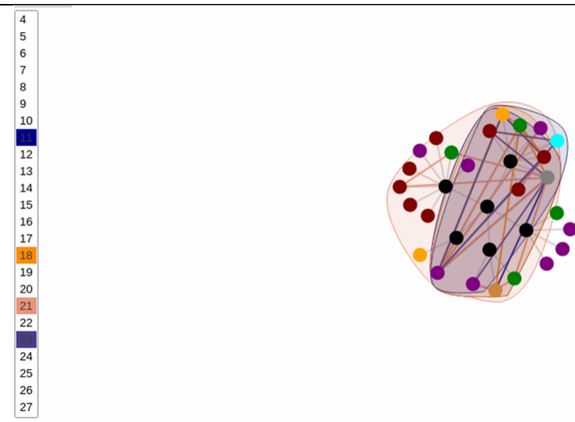
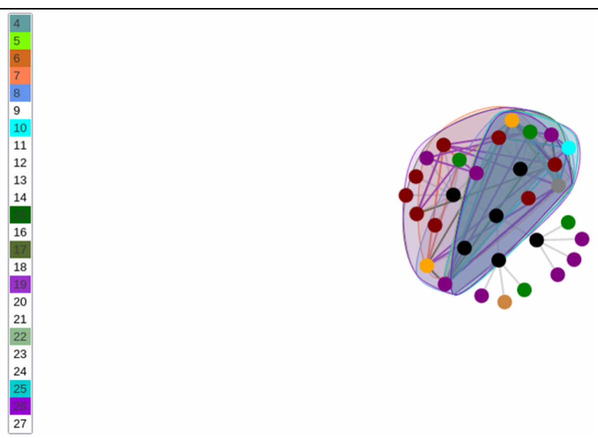
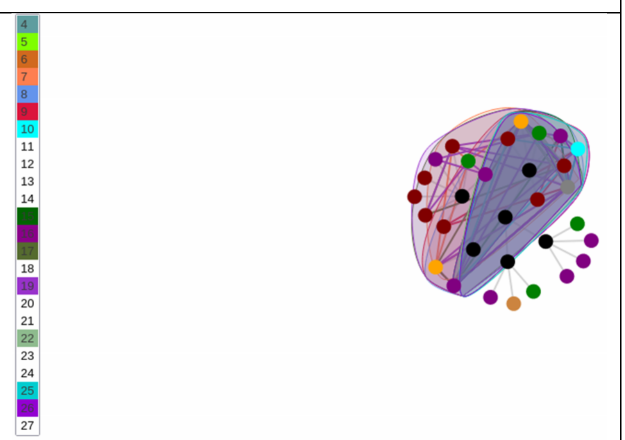
	
Blue Cluster (computer result)	*Blue Cluster (human result)*
	
Yellow Cluster (computer result)	*Yellow Cluster (human result)*
	
Green Cluster (computer result)	*Green Cluster (human result)*

FIGURE 19: CLUSTERING RESULTS COMPARED (COMPUTER VS. HUMAN)

From Figure 19 it is noticeable that students in the blue clusters seem to be interacting with the widest diversity of learning objects, while the students in the green clusters have not interacted with the learning objects in the lower right quadrant of the graph. The yellow manual cluster has students that primarily interacted with the learning objects through the center of the graph. Although there do appear to be some behaviour patterns that are common to the clusters, a visual inspection of the screenshots is not enough to know for sure if the clustering results are meaningful.

The teacher who created the clustering analysis and manual adjustments was interviewed to gather their insights into why some students were moved between clusters. The teacher indicated that they had reviewed the student's learning behaviour patterns and concluded, based on the learning object graphs, that some students belonged to other clusters than the system put them in. As such, manual adjustments were made. The teacher also mentioned that many students had limited behaviour data because the students had only a few weeks to use the Moodle course rather than a full semester. The students were also guided by the teacher as to how to use the Moodle course materials, possibly skewing the results. The interview revealed that the collected data may not be of ideal quality and may not generalize to other students or courses. The teacher indicated willingness to spend more time using Behaviour Analytics to produce more data to analyze but noted that Moodle is no longer the learning management system in use at his school, making generating more data difficult or impossible.

5.3 Findings and Discussion

The collected data is analyzed to determine whether the research questions (RQ1, RQ2, RQ3) are answered. The first hypothesis (H1) is that the clustering results done by Behaviour Analytics compares against the results adjusted by the teacher are similar. Comparing the system generated clusters with the teacher's manually adjustments based on precision, recall, and F-scores. Table 16 lists the performance measures based on the computer and manual clustering results.

TABLE 16: SYSTEM AND MANUAL CLUSTERING RESULTS COMPARED

Cluster	System	Manual	Precision	Recall	$F_{0.5}$	F_1	F_2
1	13, 18, 23, 24, 27	12, 13, 14, 20, 24, 27	0.6	0.5	0.58	0.55	0.52
2	1, 2, 3	1, 2, 3	1	1	1	1	1
3	4, 5, 6, 7, 8, 10, 15, 17, 19, 22, 25, 26	4, 5, 6, 7, 8, 9, 10, 15, 16, 17, 19, 22, 25, 26	1	0.86	0.97	0.92	0.88
4	9, 11, 12, 14, 16, 20, 21	11, 18, 21, 23	0.29	0.5	0.31	0.36	0.44
Total			0.74	0.74	0.74	0.74	0.74

If the admin data that was removed from this dataset is added back in, the totals go up to 0.75, but the admin data was removed because it was not supposed to be there, so F-score 0.74 is the result of the study. This result shows that the system results are more similar than they are dissimilar but are nowhere near perfect. The result is from a single small sample of student data from a single course. Other courses should be tested with many more students before it can be definitively declared that BA clusters students correctly about 74% of the time. Other students in other courses may produce results that are considerably different than what was obtained from the given sample.

Due to the way it works, the clustering done by the BA plugin has the potential to group students well based on their behaviour. Students who spend time interacting with similar learning objects will have similar centroids and could easily get clustered together. In the given sample, the 3 students who had minimal interaction with the course were all perfectly clustered together because of similar behaviour. A caveat to that is when one student has interacted with similar objects to another student, but not as much as the other student. These students may have similar centroids, but their behaviour varies in how often they are accessing the learning material. Another potential pitfall of the BA clustering method is that if a low interacting student happened to click only a single node in the graph, but it happened to be where higher interacting students' centroids wound up as well. In the given sample, the 3 low interacting students all clicked nodes away from where the other students' centroids gravitated towards, but this may not be the case with other courses and students.

Exploring research question RQ2 can be done by comparing the ILS questionnaire results with the computer and manual clusters in an attempt to determine if either clustering result did well clustering according to learning style and if so, which was better. Tables 17 and 18 list the performance measures based on the computer and manual clustering results against the ILS in precision, recall, and F-scores as before.

TABLE 17: SYSTEM AND ILS RESULTS COMPARED

Cluster	System	ILS	Precision	Recall	$F_{0.5}$	F_{1}	F_{2}
1	13, 18, 23, 24, 27	6, 10, 11, 16, 17, 18, 23, 24	0.6	0.38	0.54	0.46	0.41
2		4	0	0	0	0	0
3		26	0	0	0	0	0
4	4, 5, 6, 8, 10, 17, 19, 22, 25, 26	5, 8, 9, 12, 13, 19, 20, 21, 22, 25, 27	0.5	0.46	0.49	0.48	0.46
5	9, 11, 12, 14, 16, 20, 21	14	0.14	1	0.17	0.25	0.46
Total			0.41	0.41	0.41	0.41	0.41

TABLE 18: MANUAL AND ILS RESULTS COMPARED

Cluster	Manual	ILS	Precision	Recall	$F_{0.5}$	F_{1}	F_{2}
1	12, 13, 14, 20, 24, 27	5, 8, 9, 12, 13, 19, 20, 21, 22, 25, 27	0.67	0.36	0.57	0.47	0.4
2		4	0	0	0	0	0
3	4, 5, 6, 8, 9, 10, 16, 17, 19, 22, 25, 26	6, 10, 11, 16, 17, 18, 23, 24	0.33	0.5	0.46	0.4	0.45
4	11, 18, 21, 23	26	0	0	0	0	0
5		14	0	0	0	0	0
Total			0.36	0.36	0.36	0.36	0.36

It can be seen from Tables 17 and 18 that neither the computer generated clusters nor the teacher altered clusters are well aligned with the ILS results. The computer generated clusters are only slightly better than the teacher's alterations. From the given data, it seems that the BA plugin cannot predict student's learning style from their behaviour. But, different courses and students may produce different results. Using a different learning style model could also cause different results. Even if the BA plugin successfully clustered students according to learning style, then there would still need to be a mechanism to determine which cluster belongs to which learning style.

The astute observer will notice that the student IDs in Table 15 do not match those in Tables 16, 17, and 18. The reason is that the anonymous IDs in Table 15 come from the ILS results which do not include the students that did not fill out the questionnaire and the anonymous IDs in Tables 16, 17, and 18 come from the BA plugin which include all students and an administrator.

The research question RQ3 asked whether the BA clustering was better at predicting learning styles than the teacher's alterations to the clusters. From Tables 17 and 18 it appears that BA does do a better job of predicting learning styles, but not by very much. Without other datasets to analyze, it cannot be known for certain if BA would always do a better job of clustering according to learning style.

Chapter 6. Conclusion

6.1 Summary

The current research was conducted with the goal of improving the student learning experience. Students have a better learning experience when their learning material is customized to fit their learning preferences. This research set out to answer the questions: Can the log file clustering provided by the Behaviour Analytics plugin detect student learning styles? And can providing students a personalized study guide based on Behaviour Analytics improve the students' learning experience? However, data limitations forced the adoption of an alternate set of questions and hypothesis:

RQ1: Can the Behaviour Analytics plugin cluster students properly according to their actions in the Learning Management System?

H1: The clustering result done by Behaviour Analytics compares against the final results adjusted by Teacher are similar.

RQ2: Do the clustering results done by Behaviour Analytics and Teacher match the learning style found through Index of Learning Styles questionnaire?

H2.1: The clustering result done by Behaviour Analytics compares against the grouping based on responses of Learning Style Inventory questionnaire are similar.

H2.2: The clustering result adjusted by Teacher compares against the grouping based on responses of Learning Style Inventory questionnaire are similar.

RQ3: Has the Behaviour Analytics plugin done a better job than Teacher on clustering students based on their actions in the Learning Management System?

H3: The clustering result done by Behaviour Analytics compares against the grouping based on responses of Learning Style Inventory questionnaire is closer than the clustering result adjusted by Teacher.

To inform this research, a systematic literature review was done, which first looked at similar systematic literature reviews. From these, keywords were chosen to search for literature pertaining to the subjects of log data analysis, personalized learning, and learning style prediction. The most relevant literature was selected through keyword search and exclusion criteria filtering, which was then analyzed both quantitatively and qualitatively.

Finding the answer to the research question involved constructing the personalized study guide plugin. The plugin offers a personalized study guide to each student based on the student's preferences. It does so by first analyzing the relationship between learning objects in a course using natural language processing to find the similarity between learning content with the Learning Object Relation Discovery (LORD) plugin. The similarity is used to construct a Learning Object Graph (LOG), where the learning objects are nodes and the distance between nodes is based on the similarity of the learning content. The LOG is used by the Behaviour Analytics plugin that adds to the graph sequential student behaviour data as links between the learning object nodes. The plugin then computes a centroid point for each student based on their behaviour graph and uses k-means to cluster the student centroids into learning style groups.

The Personalized Study Guide plugin utilizes that learning style classification to inform its personalization process. Each different learning style group will get a different study guide that is personalized to the learning preferences of the group.

To test the personalized study guide, a group of students was to be invited to participate in this research. Those that chose to use the personalized study guide would then be asked to fill out a System Usability Scale (SUS) survey about their experience with using the plugin. However, things never got that far. The original research questions had to be redone and the PSG plugin was never used. Nor was the LORD plugin. Nor were many features of BA. The small dataset that was produced and analyzed for this research is too little to effectively determine if the BA plugin is in fact useful or not.

6.2 Limitations

The current research aims to improve the student learning experience by providing a personalized study guide to each learning style group. Some students may not fit neatly into a learning style category, so the personalization offered may not be appropriate. Ideally, each student should be considered as an individual, but it is easier to build personalization systems when the system only considers groups, so there may be some students that never derive much benefit from using the system.

Fuzzy clustering is an approach that has been used to mitigate this problem, but when the personalizations themselves are group based, the student must still fit into one group more than the others for the system to decide on the personalization used. Ultimately, each student should have a personalized tutor capable of meeting the student's individual needs. While such an ideal is not possible in real life due to human resource constraints, it might be possible with computerized tutors. The current research is well short of the ideal computerized tutor but tries to help bridge the gap between the two. The current research is also limited to testing only the usability of the personalized study guide and does not evaluate it further.

The primary limitation in this research was lack of data. Data needs to be generated so that it can be analyzed to determine if the research questions are answered. Not enough data means analysis results may not be statistically significant as there are many factors that determine the clustering that BA does. One factor is in the graph and the orientation of the nodes. The graph used was a system generated random graph that was not altered. The BA plugin has an interface to change the layout of the nodes and also remove them entirely from consideration. The LORD plugin was also supposed to be used to help generate a graph and BA has an option to generate the graph using Lag Sequence Analysis (LSA). None of these options were explored in the singular clustering analysis.

Given the number of possible graph arrangements, metrics to compare against the clustering results, student selections, and time slice selections it would require a much larger sample size than was obtained to achieve statistical significance. If a large enough sample size could be obtained to determine something about what BA clustering indicates in a single course, it might only apply to that course and the same amount of data would then need to be collected from a number of different courses. The ability to generate so many possible clustering results with a single course and a single class of students is itself a limitation. The BA clustering may indicate different metrics in a single course depending on how it is used and certainly has the potential to indicate different metrics across courses.

A limitation with PSG is that it needs a course with a variety of different learning object types. A course with many of the same type of learning object can not have the objects sorted by learning style because the objects all have the same one. A course with few if any non-standard objects is also important. The standard object types that are built into Moodle are mapped to learning styles, but third-party learning object types are not and can also not be sorted.

6.3 Future Works

Given the small sample size of the collected data and that the data was from a single course, the results are not generalizable. These same students is a different course might produce completely different behaviour patterns or clusters. Other students in the same course might produce different results and other students in other courses may also produce different results. There is simply not enough data to generalize the results found. How much data needs to be generated and analyzed to achieve generalizability? Given the number of variables going into the equation, this is a hard question to answer. There are many possible graph configurations to consider across many possible courses and many possible students with many possible time slices of data. But, without considerable amounts of data, the results found in this study may or may not generalize.

The current research used only learning styles to inform the study guide personalization. Other data could also be input into the student model to better inform the personalization. Other studies have considered a student's prior knowledge when recommending learning objects, so prior knowledge is something that could be considered for the personalized study guide as well. There are also other clustering algorithms that could be easily applied in place of the k-means, or fuzzy clustering could be used to place some students in multiple learning style clusters. Alternatively, a different learning style model could be used that might produce better results. There are a few minor alterations to the current research that could be done in the future in an attempt to improve upon the current results.

The personalized study guide in this research offered a limited analysis of student behaviour and also offered a limited personalization. The sequential analysis only considered the learning object access and did not consider the time between accesses or how the student used the different learning objects. The behaviour analysis was also restricted to a single course and did not consider student behaviour across courses. Furthermore, the personalization only involved rearranging the learning objects within the study guide and did not consider offering different types of learning objects, such as video instead of text, to different student groups. There are various alterations that could be made to the techniques used in this research that might be able to improve the results.

The sequential behaviour pattern extraction and clustering process used in Behaviour Analytics could also be applied to another type of data. In this research, learning data was used to cluster according to learning style, but it would be possible to apply the same techniques to a different type of sequential data such as customer traces from a shopping site or user interaction with a program. Alternately, different data could be used to verify the student clusters such as academic performance. Rather than evaluating whether learning styles are correlated with the clusters, student marks could be considered as potentially correlated with the clusters.

There are more features that could be added to BA plugin to make it better. It may be helpful to teachers to have the clusters generated be automatically annotated. While it is possible to see the common links graph for a cluster, an automatic annotation feature that could summarize the behaviour or common learning objects in a cluster could help teachers understand their students behaviour better.

There are also other automatic graph generation methods that could be employed to alter the graph node arrangement and change the clustering results.

Finally, the research could be finished as it was intended. As mentioned in the limitations, many of the features available with BA were not utilized and limited data was collected for analysis. Certainly, more testing of various kinds can be done with the software involved in this research.

Looking forward to the future of personalized learning, it will be interesting to see what Artificial Intelligence (AI) can do for the field. Since this research began it seems there has been an explosion of interest and ability in AI. There are Large Language Models (LLMs) available that have impressive abilities and are being used in many areas of human endeavour. Computer code is also increasingly being written by AI models, automobiles and drones are becoming increasingly autonomous. An AI powered personalized learning system is a logical application for the power of modern AI.

References

Advanced Distributed Learning Initiative (ADLI). (n.d.). Shareable Content Object Reference Model (SCORM®). Access: https://www.adlnet.gov/projects/scorm

Aeiad, E., & Meziane, F. (2019). An adaptable and personalised E-learning system applied to computer science Programmes design. Education and Information Technologies, 24(2), 1485–1509. https://doi.org/10.1007/s10639-018-9836-x

Afini Normadhi, N. B., Shuib, L., Md Nasir, H. N., Bimba, A., Idris, N., & Balakrishnan, V. (2019). Identification of personal traits in adaptive learning environment: Systematic literature review. Computers & Education, 130, 168–190. https://doi.org/10.1016/j.compedu.2018.11.005

Ahmad Uzir, N., Gašević, D., Matcha, W., Jovanović, J., & Pardo, A. (2020). Analytics of time management strategies in a flipped classroom. Journal of Computer Assisted Learning, 36(1), 70–88. https://doi.org/10.1111/jcal.12392

Aissaoui, Ouafae EL, El Madani, Y. E. A., Oughdir, L., & Allioui, Y. E. (2019a). Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. Procedia Computer Science, 148, 87–96. https://doi.org/10.1016/j.procs.2019.01.012

Aissaoui, Ouafae El, El Madani, Y. E. A., Oughdir, L., & El Allioui, Y. (2019b). A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments. Education and Information Technologies, 24(3), 1943-1959. https://doi.org/10.1007/s10639-018-9820-5

Alian, M., & Shaout, A. (2017). Predicting learners styles based on fuzzy model. Education and Information Technologies, 22(5), 2217–2234. https://doi.org/10.1007/s10639-016-9543-4

Anandhi, D., & Ahmed, M. S. I. (2019). Prediction of user's type and navigation pattern using clustering and classification algorithms. Cluster Computing, 22(S5), 10481–10490. https://doi.org/10.1007/s10586-017-1090-2

Ansari, Z. A., Sattar, S. A., & Babu, A. V. (2017). A fuzzy neural network based framework to discover user access patterns from web log data. Advances in Data Analysis and Classification, 11(3), 519–546. https://doi.org/10.1007/s11634-015-0228-4

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. Educational Technology Research and Development, 60(3), 383–398. https://doi.org/10.1007/s11423-012-9235-8

Azzi, I., Jeghal, A., Radouane, A., Yahyaouy, A., & Tairi, H. (2020). A robust classification to predict learning styles in adaptive E-learning systems. Education and Information Technologies, 25(1), 437–448. https://doi.org/10.1007/s10639-019-09956-6

Bernard, J., Chang, T.-W., Popescu, E., & Graf, S. (2017). Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithms. Expert Systems with Applications, 75, 94–108. https://doi.org/10.1016/j.eswa.2017.01.021

Cabada, R. Z., Barrón Estrada, M. L., & Reyes García, C. A. (2011). EDUCA: A web 2.0 authoring tool for developing adaptive and intelligent tutoring systems using a Kohonen network. Expert Systems with Applications, 38(8), 9522–9529. https://doi.org/10.1016/j.eswa.2011.01.145

Cassidy, S. (2004). Learning Styles: An overview of theories, models, and measures. Educational Psychology, 24(4), 419-444. https://www.tandfonline.com/doi/full/10.1080/0144341042000228834

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

Cinque, M., Della Corte, R., Moscato, V., & Sperlí, G. (2021). A graph-based approach to detect unexplained sequences in a log. Expert Systems with Applications, 171, 114556. https://doi.org/10.1016/j.eswa.2020.114556

Dani, A., & Nasser, R. (2016). Use of Intelligent Tutor in Post-Secondary Mathematics Education in the United Arab Emirates. The Turkish Online Journal of Educational Technology, 15(4), 152-162. Access: http://www.tojet.net/articles/v15i4/15414.pdf

Dorça, F. (2015). Implementation and use of Simulated Students for Test and Validation of new Adaptive Educational Systems: A Practical Insight. International Journal of Artificial Intelligence in Education, 25(3), 319–345. https://doi.org/10.1007/s40593-015-0037-0

Dorça, F. A., Lima, L. V., Fernandes, M. A., & Lopes, C. R. (2013). Comparing strategies for modeling students learning styles through reinforcement learning in adaptive and intelligent educational systems: An experimental analysis. Expert Systems with Applications, 40(6), 2092–2101. https://doi.org/10.1016/j.eswa.2012.10.014

Dung, P. Q., & Florea, A. M. (2012). A literature-based method to automatically detect learning styles in learning management systems. In: 2nd International Conference on Web Intelligence, Mining and Semantics, Craiova Romania June 13 - 15, 2012. Article No. 46. https://doi.org/10.1145/2254129.2254186

Dwivedi, P., Kant, V., & Bharadwaj, K. K. (2018). Learning path recommendation based on modified variable length genetic algorithm. Education and Information Technologies, 23(2), 819–836. https://doi.org/10.1007/s10639-017-9637-7

El Aissaoui, O., El Alami El Madani, Y., Oughdir, L., & El Allioui, Y. (2019). A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments. Education and Information Technologies, 24(3), 1943–1959. https://doi.org/10.1007/s10639-018-9820-5

Ennouamani, S., Mahani, Z., & Akharraz, L. (2020). A context-aware mobile learning system for adapting learning content and format of presentation: Design, validation and evaluation. Education and Information Technologies, 25(5), 3919–3955. https://doi.org/10.1007/s10639-020-10149-9

Fasihuddin, H., Skinner, G., & Athauda, R. (2017). Towards adaptive open learning environments: Evaluating the precision of identifying learning styles by tracking learners' behaviours. Education and Information Technologies, 22(3), 807–825. https://doi.org/10.1007/s10639-015-9458-5

Fatahi, S., Shabanali-Fami, F., & Moradi, H. (2018). An empirical study of using sequential behavior pattern mining approach to predict learning styles. Education and Information Technologies, 23(4), 1427–1445. https://doi.org/10.1007/s10639-017-9667-1

Felder, R. M. & Silverman, L. K. (1988). LEARNING AND TEACHING STYLES IN ENGINEERING EDUCATION. Engineering Education, 78(7), 674–681

Fleming, N.D. (1995). I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom. Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA), 18, 308 - 313. https://fyi.extension.wisc.edu/wateroutreach/files/2016/03/Fleming_VARK_Im_Different_Not_Dumb.pdf

Free Software Foundation (FSF). (2007) GNU General Public License, version 3. https://www.gnu.org/licenses/gpl-3.0.html

Hoppe, H. U., Doberstein, D., & Hecking, T. (2020). Using Sequence Analysis to Determine the Well-Functioning of Small Groups in Large Online Courses. International Journal of Artificial Intelligence in Education. https://doi.org/10.1007/s40593-020-00229-9

Huang, A. F. M., Wu, J. T. H., Yang, S. J. H., & Hwang, W.-Y. (2012). The success of ePortfolio-based programming learning style diagnosis: Exploring the role of a heuristic fuzzy knowledge fusion. Expert Systems with Applications, 39(10), 8698–8706. https://doi.org/10.1016/j.eswa.2012.01.212

Huang, X., Ye, Y., Xiong, L., Wang, S., & Yang, X. (2016). Clustering time-stamped data using multiple nonnegative matrices factorization. Knowledge-Based Systems, 114, 88–98. https://doi.org/10.1016/j.knosys.2016.10.007

Janning, R., Schatten, C., & Schmidt-Thieme, L. (2016). Perceived Task-Difficulty Recognition from Log-file Information for the Use in Adaptive Intelligent Tutoring Systems. International Journal of Artificial Intelligence in Education, 26(3), 855–876. https://doi.org/10.1007/s40593-016-0097-9

Karagiannis, I., & Satratzemi, M. (2018). An adaptive mechanism for Moodle based on automatic detection of learning styles. Education and Information Technologies, 23(3), 1331–1357. https://doi.org/10.1007/s10639-017-9663-5

Kassak, O., Kompan, M., & Bielikova, M. (2016). Student behavior in a web-based educational system: Exit intent prediction. Engineering Applications of Artificial Intelligence, 51, 136–149. https://doi.org/10.1016/j.engappai.2016.01.018

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

Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education, 56(3), 885–899. https://doi.org/10.1016/j.compedu.2010.11.001

Kolekar, S. V., Pai, R. M., & M. M., M. P. (2019). Rule based adaptive user interface for adaptive E-learning system. Education and Information Technologies, 24(1), 613–641. https://doi.org/10.1007/s10639-018-9788-1

Latham, A., Crockett, K., & McLean, D. (2014). An adaptation algorithm for an intelligent natural language tutoring system. Computers & Education, 71, 97–110. https://doi.org/10.1016/j.compedu.2013.09.014

Li, L. X., & Abdul Rahman, S. S. (2018). Students' learning style detection using tree augmented naive Bayes. Royal Society Open Science, 5(7), 172108. https://doi.org/10.1098/rsos.172108

Li, L.-Y., & Tsai, C.-C. (2017). Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. Computers & Education, 114, 286–297. https://doi.org/10.1016/j.compedu.2017.07.007

Li, L.-Y., & Tsai, C.-C. (2020). Students' patterns of accessing time in a text structure learning system: Relationship to individual characteristics and learning performance. Educational Technology Research and Development, 68(5), 2569–2594. https://doi.org/10.1007/s11423-020-09780-7

Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. Educational Technology Research and Development, 68(4), 1903–1929. https://doi.org/10.1007/s11423-020-09793-2

McGreal, R. (2004). Learning objects: A practical definition. International Journal of Instructional Technology and Distance Learning (IJITDL), 9(1). https://auspace.athabascau.ca/handle/2149/227

Meza-Fernández, S., & Sepúlveda-Sariego, A. (2017). Representational model on Moodle's activity: Learning styles and navigation strategies. International Journal of Educational Technology in Higher Education, 14(1), 14. https://doi.org/10.1186/s41239-017-0052-3

Niknam, M., & Thulasiraman, P. (2020). LPR: A bio-inspired intelligent learning path recommendation system based on meaningful learning theory. Education and Information Technologies, 25(5), 3797–3819. https://doi.org/10.1007/s10639-020-10133-3

Pane, J., Steiner, E., Baird, M., Hamilton, L., & Pane, J. (2017). How Does Personalized Learning Affect Student Achievement? RAND Corporation. https://doi.org/10.7249/RB9994

Poornalatha, G., & Prakash, S. R. (2013). Web sessions clustering using hybrid sequence alignment measure (HSAM). Social Network Analysis and Mining, 3(2), 257–268. https://doi.org/10.1007/s13278-012-0070-z

Ramirez-Arellano, A., Bory-Reyes, J., & Hernández-Simón, L. M. (2017). Learning Object Retrieval and Aggregation Based on Learning Styles. Journal of Educational Computing Research, 55(6), 757–788. https://doi.org/10.1177/0735633116681303

Rani, M., Nayak, R., & Vyas, O. P. (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems, 90, 33–48. https://doi.org/10.1016/j.knosys.2015.10.002

Rastegarmoghadam, M., & Ziarati, K. (2017). Improved modeling of intelligent tutoring systems using ant colony optimization. Education and Information Technologies, 22(3), 1067–1087. https://doi.org/10.1007/s10639-016-9472-2

Riestra-González, M., Paule-Ruíz, M. del P., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & Education, 163, 104108. https://doi.org/10.1016/j.compedu.2020.104108

Salehi, M., & Nakhai Kamalabadi, I. (2013). Hybrid recommendation approach for learning material based on sequential pattern of the accessed material and the learner's preference tree. Knowledge-Based Systems, 48, 57–69. https://doi.org/10.1016/j.knosys.2013.04.012

Salehi, M., Nakhai Kamalabadi, I., & Ghaznavi Ghoushchi, M. B. (2014). Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering. Education and Information Technologies, 19(4), 713–735. https://doi.org/10.1007/s10639-012-9245-5

Shi, D., Wang, T., Xing, H., & Xu, H. (2020). A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. Knowledge-Based Systems, 195, 105618. https://doi.org/10.1016/j.knosys.2020.105618

The Myers-Briggs Company. (n.d.). https://www.themyersbriggs.com/en-US/

Truong, H. M. (2016). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. Computers in Human Behavior, 55, 1185–1193. https://doi.org/10.1016/j.chb.2015.02.014

Uto, M., Miyazawa, Y., Kato, Y., Nakajima, K., & Kuwata, H. (2020). Time- and Learner-Dependent Hidden Markov Model for Writing Process Analysis Using Keystroke Log Data. International Journal of Artificial Intelligence in Education, 30(2), 271–298. https://doi.org/10.1007/s40593-019-00189-9

Valsamidis, S., Kontogiannis, S., Kazanidis, I., Theodosiou, T., & Karakos, A. (2012). A Clustering Methodology of Web Log Data for Learning Management Systems. Educational Technology & Society, 15(2), 154-167. Access: https://drive.google.com/file/d/1bOmGlB6M7ydmXV6O7cco73N-EYFvl2qy/view

VARK. (n.d.). https://vark-learn.com/

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. Computers & Education, 122, 119–135. https://doi.org/10.1016/j.compedu.2018.03.018

Wan, S., & Niu, Z. (2016). A learner oriented learning recommendation approach based on mixed concept mapping and immune algorithm. Knowledge-Based Systems, 103, 28–40. https://doi.org/10.1016/j.knosys.2016.03.022

Wu, C.-H., Chen, T.-C., Yan, Y.-H., & Lee, C.-F. (2017). Developing an adaptive e-learning system for learning excel. 2017 International Conference on Applied System Innovation (ICASI), 1973–1975. https://doi.org/10.1109/ICASI.2017.7988583

Xie, T., Zheng, Q., & Zhang, W. (2016). A behavioral sequence analyzing framework for grouping students in an e-learning system. Knowledge-Based Systems, 111, 36–50. https://doi.org/10.1016/j.knosys.2016.08.001

Appendix A

Index of Learning Styles

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.

3. When I think about what I did yesterday, I am most likely to get

    a. a picture.

    b. words.

4. I tend to

    a. understand details of a subject but may be fuzzy about its overall structure.

    b. understand the overall structure but may be fuzzy about details.

5. When I am learning something new, it helps me to

    a. talk about it.

    b. think about it.

6. If I were a teacher, I would rather teach a course

    a. that deals with facts and real life situations.

    b. that deals with ideas and theories.

7. I prefer to get new information in

    a. pictures, diagrams, graphs, or maps.

    b. written directions or verbal information.

8. Once I understand

    a. all the parts, I understand the whole thing.

    b. the whole thing, I see how the parts fit.

9. In a study group working on difficult material, I am more likely to

    a. jump in and contribute ideas.

    b. sit back and listen.

10. I find it easier

    a. to learn facts.

    b. to learn concepts.

11. In a book with lots of pictures and charts, I am likely to

    a. look over the pictures and charts carefully.

    b. focus on the written text.

12. When I solve math problems

    a. I usually work my way to the solutions one step at a time.

    b. I often just see the solutions but then have to struggle to figure out the steps to get to them.

13. In classes I have taken

    a. I have usually gotten to know many of the students.

    b. I have rarely gotten to know many of the students.

14. In reading nonfiction, I prefer

    a. something that teaches me new facts or tells me how to do something.

    b. something that gives me new ideas to think about.

15. I like teachers

    a. who put a lot of diagrams on the board.

    b. who spend a lot of time explaining.

16. When I'm analyzing a story or a novel

    a. I think of the incidents and try to put them together to figure out the themes.

    b. I just know what the themes are when I finish reading and then I have to go back and find the incidents that demonstrate them.

17. When I start a homework problem, I am more likely to

    a. start working on the solution immediately.

    b. try to fully understand the problem first.

18. I prefer the idea of

    a. certainty.

    b. theory.

19. I remember best

    a. what I see.

    b. what I hear

20. It is more important to me that an instructor

    a. lay out the material in clear sequential steps.

    b. give me an overall picture and relate the material to other subjects.

21. I prefer to study

    a. in a study group.

    b. alone.

22. I am more likely to be considered

    a. careful about the details of my work.

    b. creative about how to do my work.

23. When I get directions to a new place, I prefer

    a. a map.

    b. written instructions.

24. I learn

    a. at a fairly regular pace. If I study hard, I'll "get it."

    b. in fits and starts. I'll be totally confused and then suddenly it all "clicks."

25. I would rather first

    a. try things out.

    b. think about how I'm going to do it.

26. When I am reading for enjoyment, I like writers to

    a. clearly say what they mean.

    b. say things in creative, interesting ways.

27. When I see a diagram or sketch in class, I am most likely to remember

    a. the picture.

    b. what the instructor said about it.

28. When considering a body of information, I am more likely to

    a. focus on details and miss the big picture.

    b. try to understand the big picture before getting into the details.

29. I more easily remember

    a. something I have done.

    b. something I have thought a lot about.

30. When I have to perform a task, I prefer to

    a. master one way of doing it.

    b. come up with new ways of doing it.

31. When someone is showing me data, I prefer

    a. charts or graphs.

    b. text summarizing the results.

32. When writing a paper, I am more likely to

    a. work on (think about or write) the beginning of the paper and progress forward.

    b. work on (think about or write) different parts of the paper and then order them.

33. When I have to work on a group project, I first want to

    a. have "group brainstorming" where everyone contributes ideas.

    b. brainstorm individually and then come together as a group to compare ideas.

34. I consider it higher praise to call someone

    a. sensible.

    b. imaginative.

35. When I meet people at a party, I am more likely to remember

    a. what they looked like.

    b. what they said about themselves.

36. When I am learning a new subject, I prefer to

    a. stay focused on that subject, learning as much about it as I can.

    b. try to make connections between that subject and related subjects.

37. I am more likely to be considered

    a. outgoing.

    b. reserved.

38. I prefer courses that emphasize

    a. concrete material (facts, data).

    b. abstract material (concepts, theories).

39. For entertainment, I would rather

    a. watch television.

    b. read a book.

40. Some teachers start their lectures with an outline of what they will cover. Such outlines are

    a. somewhat helpful to me.

    b. very helpful to me.

41. The idea of doing homework in groups, with one grade for the entire group,

    a. appeals to me.

    b. does not appeal to me.

42. When I am doing long calculations,

    a. I tend to repeat all my steps and check my work carefully.

    b. I find checking my work tiresome and have to force myself to do it.

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.

Appendix B

Athabasca University
RESEARCH CENTRE

CERTIFICATION OF ETHICAL APPROVAL - RENEWAL

The Athabasca University Research Ethics Board (REB) has reviewed and approved the research project noted below. The REB is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and Athabasca University Policy and Procedures.

Ethics File No.: **24572**

Principal Investigator:
Mr. Theodore Krahn, Graduate Student
Faculty of Science & Technology\Master of Science in Information Systems (MScIS)

Supervisor/Project Team:
Dr. Maiga Chang (Supervisor)

Project Title:
Graph-based Clustering Method - Using Behaviour Analytics and LORD Moodle Plug-ins

Effective Date: November 9, 2022 **Expiry Date**: November 08, 2023

Restrictions:

Any modification/amendment to the approved research must be submitted to the AUREB for approval prior to proceeding.

Any adverse event or incidental findings must be reported to the AUREB as soon as possible, for review.

Ethical approval is valid *for a period of one year*. An annual request for renewal must be submitted and approved by the above expiry date if a project is ongoing beyond one year.

An Ethics Final Report must be submitted when the research is complete *(i.e. all participant contact and data collection is concluded, no follow-up with participants is anticipated and findings have been made available/provided to participants (if applicable))* or the research is terminated.

Approved by: **Date: November 09, 2022**

Paul Jerry, Chair
Athabasca University Research Ethics Board

Athabasca University Research Ethics Board
University Research Services Office
1 University Drive, Athabasca AB Canada T9S 3A3
E-mail rebsec@athabascau.ca
Telephone: 780.213.2033

Athabasca University

CERTIFICATION OF ETHICAL APPROVAL - RENEWAL

The Athabasca University Research Ethics Board (REB) has reviewed and approved the research project noted below. The REB is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and Athabasca University Policy and Procedures.

Ethics File No.: **24572**

Principal Investigator:
Mr. Theodore Krahn, Graduate Student
Faculty of Science & Technology\Master of Science in Information Systems (MScIS)

Supervisor:
Dr. Maiga Chang (Supervisor)

Project Title:
Graph-based Clustering Method - Using Behaviour Analytics and LORD Moodle Plug-ins

Effective Date: November 8, 2023 **Expiry Date**: November 08, 2024

Restrictions:

Any modification/amendment to the approved research must be submitted to the AUREB for approval prior to proceeding.

Any adverse event or incidental findings must be reported to the AUREB as soon as possible, for review.

Ethical approval is valid *for a period of one year*. An annual request for renewal must be submitted and approved by the above expiry date if a project is ongoing beyond one year.

An Ethics Final Report must be submitted when the research is complete *(i.e. all participant contact and data collection is concluded, no follow-up with participants is anticipated and findings have been made available/provided to participants (if applicable))* or the research is terminated.

Approved by: **Date: October 13, 2023**

Paul Jerry, Chair
Athabasca University Research Ethics Board

Athabasca University Research Ethics Board
University Research Services Office
1 University Drive, Athabasca AB Canada T9S 3A3
E-mail rebsec@athabascau.ca
Telephone: 780.213.2033

Athabasca University
RESEARCH CENTRE

CERTIFICATION OF ETHICAL APPROVAL

The Athabasca University Research Ethics Board (REB) has reviewed and approved the research project noted below. The REB is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and Athabasca University Policy and Procedures.

Ethics File No.: **24848**

Principal Investigator:
Mr. Theodore Krahn, Graduate Student
Faculty of Science & Technology\Master of Science in Information Systems (MScIS)

Supervisor/Project Team:
Dr. Maiga Chang (Supervisor)

Project Title:
Evaluating the Personalised Study Guide Moodle plugin

Effective Date: July 26, 2022 **Expiry Date**: July 25, 2023

Restrictions:

Any modification/amendment to the approved research must be submitted to the AUREB for approval prior to proceeding.

Any adverse event or incidental findings must be reported to the AUREB as soon as possible, for review.

Ethical approval is valid *for a period of one year*. An annual request for renewal must be submitted and approved by the above expiry date if a project is ongoing beyond one year.

An Ethics Final Report must be submitted when the research is complete *(i.e. all participant contact and data collection is concluded, no follow-up with participants is anticipated and findings have been made available/provided to participants (if applicable))* or the research is terminated.

Approved by: **Date: July 26, 2022**

Oscar Lin, Chair
School of Computing & Information Systems, Departmental Ethics Review Committee

---

Athabasca University Research Ethics Board
University Research Services Office
1 University Drive, Athabasca AB Canada T9S 3A3
E-mail rebsec@athabascau.ca
Telephone: 780.213.2033

Athabasca University
RESEARCH CENTRE

CERTIFICATION OF ETHICAL APPROVAL - RENEWAL

The Athabasca University Research Ethics Board (REB) has reviewed and approved the research project noted below. The REB is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and Athabasca University Policy and Procedures.

Ethics File No.: **24848**

Principal Investigator:
Mr. Theodore Krahn, Graduate Student
Faculty of Science & Technology\Master of Science in Information Systems (MScIS)

Supervisor/Project Team:
Dr. Maiga Chang (Supervisor)

Project Title:
Evaluating the Personalised Study Guide Moodle plugin

Effective Date: July 25, 2023 **Expiry Date**: July 25, 2024

Restrictions:

Any modification/amendment to the approved research must be submitted to the AUREB for approval prior to proceeding.

Any adverse event or incidental findings must be reported to the AUREB as soon as possible, for review.

Ethical approval is valid *for a period of one year*. An annual request for renewal must be submitted and approved by the above expiry date if a project is ongoing beyond one year.

An Ethics Final Report must be submitted when the research is complete *(i.e. all participant contact and data collection is concluded, no follow-up with participants is anticipated and findings have been made available/provided to participants (if applicable))* or the research is terminated.

Approved by: **Date: June 26, 2023**

Paul Jerry, Chair
Athabasca University Research Ethics Board

Athabasca University Research Ethics Board
University Research Services Office
1 University Drive, Athabasca AB Canada T9S 3A3
E-mail rebsec@athabascau.ca
Telephone: 780.213.2033

Athabasca University
RESEARCH CENTRE

CERTIFICATION OF ETHICAL APPROVAL

The Athabasca University Research Ethics Board (REB) has reviewed and approved the research project noted below. The REB is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and Athabasca University Policy and Procedures.

Ethics File No.: **24572**

Principal Investigator:
Mr. Theodore Krahn, Graduate Student
Faculty of Science & Technology\Master of Science in Information Systems (MScIS)

Supervisor:
Dr. Maiga Chang (Supervisor)

Project Title:
Graph-based Clustering Method - Using Behaviour Analytics and LORD Moodle Plug-ins

Effective Date: December 09, 2021 **Expiry Date**: December 08, 2022

Restrictions:

Any modification or amendment to the approved research must be submitted to the AUREB for approval.

Ethical approval is valid *for a period of one year*. An annual request for renewal must be submitted and approved by the above expiry date if a project is ongoing beyond one year.

A Project Completion (Final) Report must be submitted when the research is complete *(i.e. all participant contact and data collection is concluded, no follow-up with participants is anticipated and findings have been made available/provided to participants (if applicable))* or the research is terminated.

Approved by: **Date: December 09, 2021**

Jon Dron, Chair
School of Computing & Information Systems, Departmental Ethics Review Committee

---

Athabasca University Research Ethics Board
University Research Services, Research Centre
1 University Drive, Athabasca AB Canada T9S 3A3
E-mail rebsec@athabascau.ca
Telephone: 780.213.2033