IRRODL

THE INTERNATIONAL REVIEW OF RESEARCH IN OPEN AND DISTRIBUTED LEARNING

Vol 23 No 1

February 2022

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 23, Number 1

February – 2022

Editorial – Volume 23, Issue 1
Special Issue: AI e-Learning and Online Curriculum

Ting-Chia Hsu[1], Hal Abelson[2], and Natalie Lao[3]
[1]Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan; [2]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, USA; [3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA

This special issue was created due to the growing impact of artificial intelligence (AI) in education. As such, we collected seven articles for this issue that captured this interesting phenomenon. Each article underwent a double-blind peer review to ensure that only the most relevant and highest quality papers made this special issue.

The trend of AI in online-learning research is important. From **Hwang**, **Tu**, and **Tang**'s review study, we can see that the online learning system interactions, which can facilitate personalized, adaptive, and collaborative learning, became a focus in recent years. Regarding implementation, advancements in hardware processing speeds, networking, and machine learning algorithms are assisting in the advancement of teaching and learning. For example, AI techniques can be used to predict failure in online learning, the detection of a student's risk of dropping out, or the prediction of student course satisfaction scores. **Tzeng**, **Lee**, **Huang**, **Huang**, and **Lai**'s empirical study revealed that their early warning system, using the fifth-week model for online learning, successfully predicted student performance with an accuracy exceeding 83%. **Rodríguez**, **Guerrero-Roldán**, **Baneres,** and **Karadeniz** employed a nudging intervention mechanism based on the AI technique into positively affecting the learners' performance and dropout reduction in online learning. However, Tzeng, Lee, Huang, Huang, and Lai used deep learning to assess the student experience with MOOCs and constructed a deep learning model to accurately predict student satisfaction without relying on questionnaire responses.

The global COVD-19 pandemic is forcing more and more schools to experiment with online education. This puts a premium on curricula and tools that can lead to effective distance learning and also raises a host of additional considerations for educational technology research and practice. In **Araka**, **Oboko**, **Maina**, and **Gitonga**'s review of Educational Data Mining (EDM) techniques and how they have been widely applied in online learning, they discovered an optimal EDM algorithm, Agglomerative Hierarchical, for identifying the levels of self-regulated learning profiles in online learning environments. To connect information by using AI techniques is also valuable for personalized learning. To overcome the huge amount of information in search engines, **Cheng**, **Cheng,** and **Huang** developed an Internet articles retrieval agent combined with dynamic associative concept maps, in order to help students find intended articles in their searches.

Regarding curriculum, the growing influence of AI on society makes it important for students to understand the fundamentals of AI and the application and use of AI in daily life. Cloud computing, open databases, and everything from application programming interfaces to large machine-learning services are making it possible for even young children to experiment with applications that only five years ago were only accessible to certain researchers. For example, the conversational AI learning platform was developed for MIT's App Inventor in 2019. **Hsu**, **Abelson,** and **Van Brummelen** employed this instructional tool to a junior high school in Taiwan and the results showed that the interactive learning tool for programming the AI application contributed to an enhancement of women in technology. The learning effectiveness of female students was significantly better than that of male students regardless of whether they used an experiential learning approach or a conventional approach. These curricula and tools can have enormous worldwide influence when they are shared open-access and worldwide so that everyone can build on and improve them.

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 23, Number 1

February – 2022

Using Few-Shot Learning Materials of Multiple SPOCs to Develop Early Warning Systems to Detect Students at Risk

Yung-Hsiang Hu
National Yunlin University of Science and Technology

Abstract

Early warning systems (EWSs) have been successfully used in online classes, especially in massive open online courses, where it is nearly impossible for students to interact face-to-face with their teachers. Although teachers in higher education institutions typically have smaller class sizes, they also face the challenge of being unable to have direct contact with their students during distance teaching. In this research, we examined the online learning trajectories of students participating in four small private online courses that were all taught by one teacher. We collected relevant data of 1,307 students from the campus learning management system. Subsequently, we constructed 18 prediction models, one for each week of the course, to develop an EWS for identifying students in online asynchronous learning at risk of failing (i.e., students who fail their final examination). Our results indicated that the fifth-week model successfully predicted student performance, with an accuracy exceeding 83% from the eighth week onward.

Keywords: precision education, SPOC, early warning system, portability of prediction model, LMS

Introduction

Learning management systems (LMSs) are used to quantify the learning behavior of students, enabling teachers to obtain data that are unavailable through face-to-face teaching in physical classrooms. Teachers can model or predict students' behaviors by using data mining or analysis (Papamitsiou & Economides, 2014). Massive open online courses (MOOCs) are particularly suitable for learning analytics or building prediction models because they involve the accumulation of large amounts of student data, which is helpful for the early detection of students who may be unable to complete such an online course (He et al., 2015) or for predicting academic results (Li et al., 2017). An early warning system (EWS) for online teaching is a precision teaching tool. Institutions of higher education have achieved digital transformation through the value-added application of learning data. Teachers have consequently become adept at running online courses, which may include setting up decision support systems (Kotsiantis, 2011), conducting instructional interventions at the most appropriate time by using EWSs (Howard et al., 2018), and predicting academic failure (Costa et al., 2017). Research in this field has focused on collecting data on students who are "at risk" or "off track" and determining why they failed or ceased learning; however, studies have focused on the period following the completion of courses, which is too late to provide adequate support to these students (Hu et al., 2014). Related research has also revealed that teachers can use the LMS data of single online courses on platforms such as Moodle (Cerezo et al., 2016; Romero et al., 2008) and Blackboard (Morris et al., 2005; Tempelaar et al., 2015) to build effective predictive models as warning systems (Hu et al., 2014; Macfadyen & Dawson, 2010). However, for small private online courses (SPOCs) in universities, current empirical research has focused on how the demand for teachers to build EWSs for asynchronous distance teaching courses through small samples from SPOCs has decreased. This decline may be due to the limitations of having fewer students in a class or the convenience of face-to-face consultations between teachers and students on campus. Our review of the literature also revealed that few teachers are able to use the data from multiple courses in the LMS of their institution to successfully develop portable prediction models as warning systems. Researchers have argued that this may be due to the differences between the courses and their instructional design (Gašević et al., 2016; Macfadyen & Dawson, 2010). Even if the data learning models of various courses within an institution are designed with high prediction accuracy, substantial differences may remain in the accuracy of the models (Conijn et al., 2017; Gašević et al., 2016). Therefore, although more institutions of higher education are offering SPOCs, research analyzing the use of few-shot learning materials for developing warning systems for SPOCs remains limited. Using few-shot learning is to predict something based on a few limited training examples. Currently, teachers are facing 10%–20% higher dropout rates for online courses than for face-to-face courses (Bawa, 2016). Thus, teachers require tools to help them identify struggling students before they drop out or fail. In this study, we collected small-sample data from different courses taught by the same teacher, while the courses were running, to build a portable student learning prediction model that can act as a warning system in SPOCs. Students who are at risk can be identified by analyzing data from the students' online learning trajectory that are accumulated and entered weekly into the LMS. "Students at risk" in the current paper refers to students who scored lower than 60 points on the course's final assessment. In this study, we addressed the primary research question—How can teachers use few-shot learning materials from multiple SPOCs to develop an EWS to detect students at risk?—as well as the following two related research questions:

1. How far in advance can the model predict a student's academic performance?

2. Can the model be used to predict academic performance in other courses taught by the same teacher?

Educational Data Mining

Data mining is widely used in educational institutions. The goal of educational data mining (EDM) is generally to explore the meaning behind data to improve the teaching process (Saa et al., 2019). In EDM, statistical models, mathematical algorithms, and machine learning methods are employed to analyze large data sets and reveal the correlation between learning behavior patterns and results. EDM enables teachers to gain an overview of the effective learning and behavior of students in the learning process (Ramaswami & Bhaskaran, 2009). Baradwaj and Pal (2011) summarized common data mining algorithms, including classification, clustering, the regression technique, the association rule, neural networks, decision trees, and the nearest neighbor method. Numerous researchers have applied these EDM techniques to predict student performance (Francis & Babu, 2019; Okubo et al., 2017; Sana et al., 2019).

EDM involves several steps. The first step is to determine the purpose of the research and collect data from an appropriate educational environment. The second step is to perform data preprocessing procedures. Subsequently, a prediction model is trained. After the model or pattern is established, the EDM results can provide the teacher with feedback for decision making or intervention. EDM has several applications such as predicting student performance; providing feedback for supporting instructors; offering personalization or recommendations to students; creating alerts for stakeholders; and performing student modeling, domain modeling, and student grouping and profiling (Baker et al., 2012; Romero & Ventura, 2013).

Along with the popularization of distance education, EDM research on LMS databases has also increased. For example, Chen et al. (2018) analyzed students' learning behavior data in short online courses and predicted students' learning performance at an early stage, i.e., after the first week of class (area under the curve ≥ 0.7). Kim et al. (2018) used deep learning to predict the results of students enrolled in online courses. Another study analyzed the LMS data of 658 students from nine courses in the first week and found that the online learning behaviors of students who passed the course differed significantly from those of students who did not pass (Milne et al., 2012).

As mentioned, EDM can be used to predict student learning performance, which then enables teachers to intervene early to improve student learning effectiveness. Currently, teachers can apply EDM technology first to establish a predictive model and subsequently to determine students' actual behavior in the LMS; teachers can then apply a data-driven teaching intervention. This process involves teachers establishing a scientific EWS to help students succeed.

EWSs in Education

EWSs have been used by educational institutions to identify students who are at risk or off track (Barry & Reschly, 2012). An EWS helps teachers understand students' behavior and performance through the collection of student behavioral data and building of a prediction model based on an algorithm. For example, researchers analyzed the behavioral data of students in distance courses at the Open University in the United Kingdom to predict their participation rate (Hussain et al., 2018). Teachers of distance courses can improve their students' learning and participation by establishing monitoring and guidance strategies on the basis of information from an EWS (Rodrigues et al., 2016) and providing timely interventions and remedies, especially in situations where a student is unable to satisfy specific indicators (Howard et al., 2018). One of Europe's largest distance education institutions, the Open University, developed four prediction models to identify students at risk of failure at an early stage of a course; these results are provided to teachers every week in the form of a feedback dashboard (Wolff et al., 2014).

Baker et al. (2015) built a model to make early predictions regarding the success and failure of students by analyzing students' online course activity data. The accuracy rate of the model in identifying students most likely to perform poorly was 59.5%. Other research used the EWS plug-in on Moodle to build prediction models, and the accuracy rate was 60.8% (Jokhan et al., 2018). The model developed by Conijn et al. (2016) for predicting whether students would be able to pass their courses achieved an overall accuracy rate of 68.7%. Related research has revealed that EWS prediction models differ in terms of their accuracy in various distance courses. However, the key to a successful EWS remains whether teachers are able to obtain a highly accurate prediction model.

To enable the wider use of prediction models, researchers have considered the portability of such models (Gašević et al., 2016; Jayaprakash et al., 2014). For example, in the 2011 Open Academic Analytics Initiative, an open-source model for predicting student success was developed (Lauría et al., 2012). Subsequently, these researchers performed a cross-institution practical test with data from Purdue University and Marist College (N = 18,968 and 27,276, respectively) to assess the portability of the student performance prediction model. The results revealed that although the LMS as well as teaching methods and types differed between these two institutions, similarities could be found in the student performance prediction model and related analysis. Another study investigated the portability of prediction models among various courses in the same institution, revealing poorer results than those obtained in the aforementioned research. The researchers suggested that the poor results were due to the difference in instructional design between the courses (Rienties et al., 2015). Thus, if highly dissimilar instructional designs are used in different courses, considerable disparities might also appear in the degree of use of the LMS module.

To enable regular teachers to use small samples from multiple SPOCs to promote precision education, scholars have expanded empirical research to consider the portability of prediction models. In the current research, we collected small-sample data from four asynchronous distance courses offered through an LMS at a public university of science and technology in central Taiwan; the courses were all taught by the same teacher. The data were used to build a prediction model that was then developed into an EWS for identifying students at risk of failing the course; the EWS was subsequently tested on a new course. Because the courses were all taught by the same teacher, their instructional designs were highly similar. This mitigated the effect of instructional design differences on the model.

Methodology

Participants and Data Collection

The LMS used in this research recorded every student's detailed learning activities in a database, including platform logins; page clicks; test completions; the opening, closing, and downloading of course materials; the upload of assignments; assignment grades; and browsing and posting behavior in the discussion area. Data on student activities were saved in a log file format, which meant that a record would be generated whenever an activity occurred. We used an application programming interface (API) to gather the necessary information for the prediction and analysis model. We collected a total of 354,668 logs from the second semester of the 2017 academic year (2017–2018) and first semester of the 2018 academic year (2018–2019). These courses were all asynchronous online courses with a total of 1,278 students. The courses and their assignments were designed in accordance with the Taiwanese Ministry of Education's digital course certification. Although the courses and their content differed, they were similar in their instructional design and course requirements, such as the weighting of grades, examinations, discussion topics, and number of assignments. Each asynchronous online course lasted 18 weeks. A summary of the online instructional design is provided in Table 1.

Table 1

Instructional Design of the Courses

Week	Activity
1, 12	Synchronous teaching–Introduction, keynote speech
2~8, 10-11, & 13-17	Asynchronous teaching (video)
	Asynchronous discussion (forum) *9
	Quiz *9
	Assignment *2
9, 18	Midterm/Final online exam

Note. The asterisk (*) above means frequency of learning activities.

Data Preprocessing

Data preprocessing, including data integration and data aggregation, was conducted on data from the LMS database to build the prediction model. The preprocessing stage of this research involved four steps. The first step was to filter out possible features from the database. We used analysis of variance as the basis for filtering learning features. We used the R 3.6.3 data mining software. Twenty features were generated (Table 2).

Table 2

Description of Features Used in the Prediction Models

Feature no.	Name	Description
1	view_link_count	No. of views of supplementary materials
2	create	No. of articles posted in the discussion area
3	like	No. of likes for articles posted in the discussion area
4	read	No. of articles viewed in the discussion area
5	online_video_count	No. of clicks on teaching videos
6	forum_count	No. of clicks on the discussion area webpage
7	online_video_time	Time spent on the teaching videos webpage
8	total_mobile_time	Time spent using mobile devices to access the platform
9	weekday_time	Teaching video viewing duration (Monday–Friday)
10	weekend_time	Teaching video viewing duration (Saturday–Sunday)
11	morning_time	Teaching video viewing duration (morning)
12	noon_time	Teaching video viewing duration (afternoon)
13	night_time	Teaching video viewing duration (night)
14	total_watch_time	Total teaching video viewing duration
15	download_count	Number of downloads
16	homework_count	Number of times the assignment was clicked
17	homework_time	Assignment browsing duration
18	forum_time	Forum browsing duration
19	total_non_mobile_time	Time spent using computer equipment to access the platform
20	total_use_time	Total time accessing the course platform

The focus of this research's prediction model was on predicting whether a student would be able to pass the final examination. Every student was assigned a specific label, namely *pass* or *fail*. If the student obtained a score of ≥60% for the examination, they received the pass label; otherwise, the fail label was applied. We collected the information of 1,278 students, among which 1,135 passed and 143 failed.

The second step of the preprocessing stage was to collate statistical information that represented every week's cumulative learning progress. We gathered this cumulative learning progress information because the distance courses were all asynchronous. The teacher allowed the students to set their own speed for completing the online learning task within the 18 weeks of the semester. Subsequently, because this research used an unsupervised learning algorithm, an autoencoder was set up. Therefore, the third step involved using [0,1] normalization to normalize the characteristic variables; that is, the range of the characteristics was converted to the 0–1 range. In addition to the data used for classification (i.e., academic performance), all other variables were also normalized. The calculation is expressed in Formula (1).

$$X_{nom} = \frac{X - X_{min}}{X_{max} - X_{min}} \epsilon[0,1] \tag{1}$$

The fourth step was to divide the information into a training set and test set. We randomly split the information into the training set and test set at a ratio of 7:3. The information in the training set was used to train the model, and the test set was used to evaluate the model to prevent the model from displaying over-fitting results.

Building the Model

Machine learning involves the automatic identification of a complex pattern according to the features extracted from a given data set and the making of an intelligent decision regarding new data (Kotsiantis et al., 2004). We employed a convolutional neural network (CNN) to build the prediction model.

We designed the prediction and analysis model in Python (Bowles, 2015) and used the PyTorch deep learning framework. A total of 18 predictive models were obtained in this research. Each forecasting model was based on 1 week (7 days) of data. When selecting training samples for the weekly predictive model, we selected the data set of students who had actual learning records in the LMS that week. Students who did not exhibit learning behavior that week were excluded from the training model sample for that week. To verify the model, we only selected 70% of each week's student samples for each week's model training. The remaining 30% was retained as the test data set of the predictive models.

Finally, to verify the portability of the prediction models, we gathered data from the Introduction to Artificial Intelligence distance course (N = 59) from the 2019–2020 academic year. That course was selected for verifying predictive models because it was taught by the same teacher and included a similar teaching design and similar course requirements as the courses used for the training models. Moreover, the course was offered at the same institution and used the same LMS as the other four courses.

CNN Performance Evaluation

We used a confusion matrix to verify the prediction model performance in classification. The confusion matrix is a binary classification, which is displayed in a two-by-two table. This table shows the training and performance of the network. The confusion matrix for each week is listed separately, and its format is presented in Table 3.

Table 3

Confusion Matrix for Binary Classification

		Actual	
		Passed	Failed
Predicted	Passed	TP	FP
	Failed	FN	TN

Note. TP = true passed; FP = false passed; FN = false negative; TN = true negative.

True passed (TP) indicates the student was predicted to pass and eventually did pass. True negative (TN) reveals the number of failing students who were classified accurately. False passed (FP) refers to the number of students who failed the course but had been predicted to pass. False negative (FN) denotes students who were predicted to fail but eventually passed.

The accuracy, sensitivity, specificity, and precision values were calculated from the confusion matrix (Saito & Rehmsmeier, 2015). The relevant values for each model were calculated using equations 2 to 5.

$$Accuracy(ACC) = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Sensitivity(Recall, TPR) = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity(SP, TNR) = \frac{TN}{TN + FP} \tag{4}$$

$$Precision(PPV) = \frac{TP}{TP + FP} \tag{5}$$

The F_β measure (*F* score) was obtained using the precision and sensitivity (recall) values (Toraman et al., 2019). A β value of 0.5, 1, or 2 is typically used (Goutte & Gaussier, 2005). Equation 6 was used to obtain the *F* score. In this study, β was 2.

$$F - measure = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{6}$$

A commonly used metric when performing classification is accuracy (Hanley & McNeil, 1982; He & Garcia, 2009). Precision is equivalent to the positive predictive value, and specificity is equal to 1; the TPR(true positive rate) and sensitivity are equivalent to the recall rate, respectively.

Results

Descriptive Statistics and Data Preprocessing

We selected four courses for creating predictive models and one course for verifying the portability of the predictive models. The descriptive statistics are shown in Table 4.

Table 4

Descriptive Statistics of Courses

Characteristic	Course name			
	Data Science (I) (n = 306)	Data Science (II) (n = 355)	Digital Social Innovation (n = 313)	Psychology (n = 304)
School year	2017–2018	2017–2018	2018–2019	2018–2019
College				
Humanities	40	59	29	38
Engineering	116	100	137	166
Management	74	95	79	50
Design	76	47	68	50
Sex				
Female	146	164	139	138
Male	160	191	174	166
Year of study				
1	59	44	80	85
2	100	133	49	80
3	95	150	66	122
4	47	27	114	15
Extension	5	1	4	2

Table 5 presents the descriptive statistics of every feature and the result of the test for statistical differences.

Table 5

Descriptive Statistics of Features

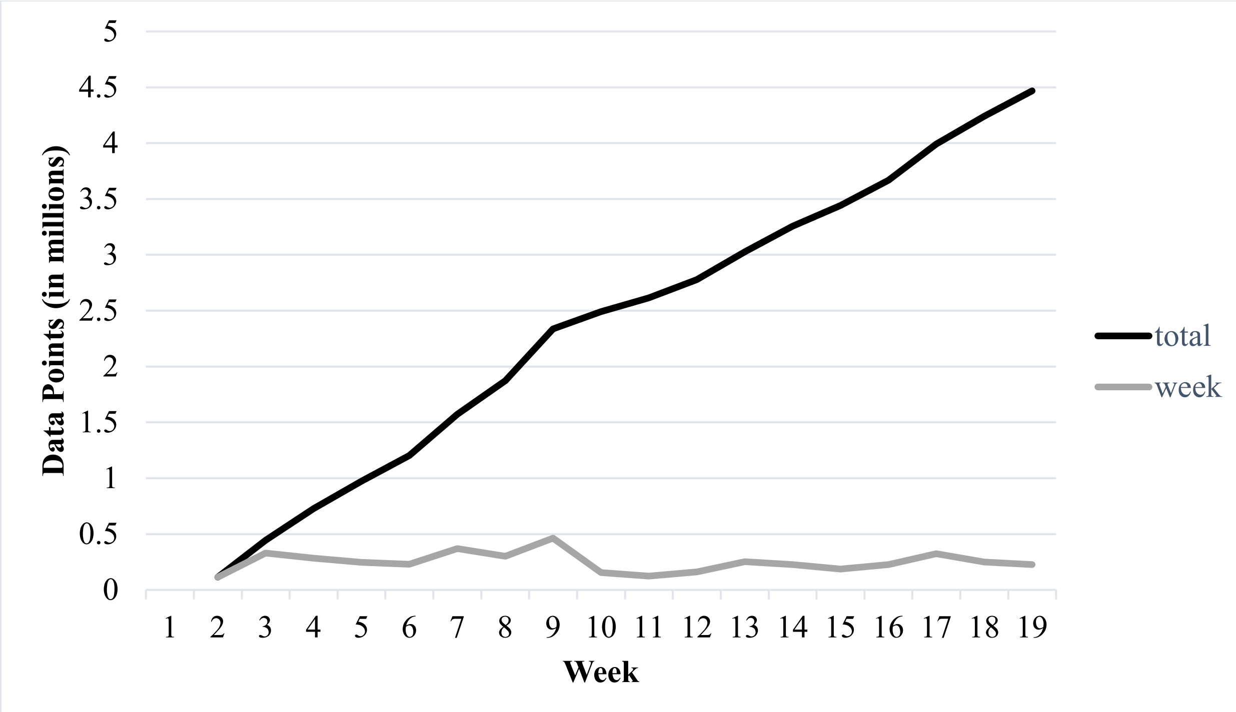
Features' column name	Passed		Failed		*t* Test
	M	*SD*	*M*	*SD*	t-value
download_count	19.8	17.4	7.3	11.2	-14.40*
view_link_count	12.4	12.5	4.5	8.0	-12.79*
create	8.6	11.8	4.3	9.0	-6.59*
like	16.0	19.3	5.7	10.3	-11.96*
read	1037.7	988.3	411.4	728.4	-11.70*
homework_count	49.6	51.6	23.8	40.6	-8.82*
online_video_count	113.9	63.8	50.4	55.3	-16.39*
forum_count	156.2	127.7	74.3	111.6	-10.50*
homework_time	6861.0	10562.8	5097.1	33819.3	-.85**
online_video_time	52810.8	74392.0	19381.6	47743.4	-9.07*
forum_time	28565.9	48707.7	14588.9	52455.8	-3.99*
total_mobile_time	8974.1	12334.2	3928.8	7031.2	-8.84*
total_non_mobile_time	113422.0	117571.7	53263.1	142109.9	-6.46*
total_use_time	122396.1	117557.1	57191.9	142438.7	-6.99*
weekday_time	32475.2	22787.0	12932.3	18103.1	-15.04*
holiday_time	12039.6	12283.3	5833.7	10229.3	-8.56*
morning_time	11825.2	11611.4	5679.2	9516.3	-9.08*
noon_time	21604.5	16790.5	8850.8	12278.8	-14.10*
night_time	11085.1	10550.1	4236.1	6632.2	-13.27*
total_watch_time	44514.8	27379.6	18766.0	22748.2	-15.97*

$^*p < .001$, $^{**}p = .39$.

Figure 1 displays the number of data points accumulated per week for all four courses. The total number of data points was 4,468,906.

Figure 1

Cumulative Data Points and Weekly Distribution of the Four Courses



Prediction Model

To create an early-stage prediction model, we obtained data on the features from the training set each week. We created a total of 18 prediction models based on each week's accumulated data. The confusion matrix was used to determine the specificity, precision, sensitivity, F-Measure and accuracy of the models. The results presented in Table 6. indicated that when looking at accuracy column, we found that the average percentage ranges from 59% at the 2nd week to 84% at the 18th week in training our model. However, the percentage ranges from 57% at the 7th week to 84% at the 18th week in testing our model. Notably, the accuracy of training data rises from 59% at the 7th week to 80% at the 8th week and the accuracy of testing data rises from 57% at the 7th week to 77% at the 8th week. Altogether, it suggests that we could predict whether students will fail or not in the middle of 18 weeks.

Table 6

Specificity, Precision, Sensitivity, F-Measure, and Accuracy Results (%)

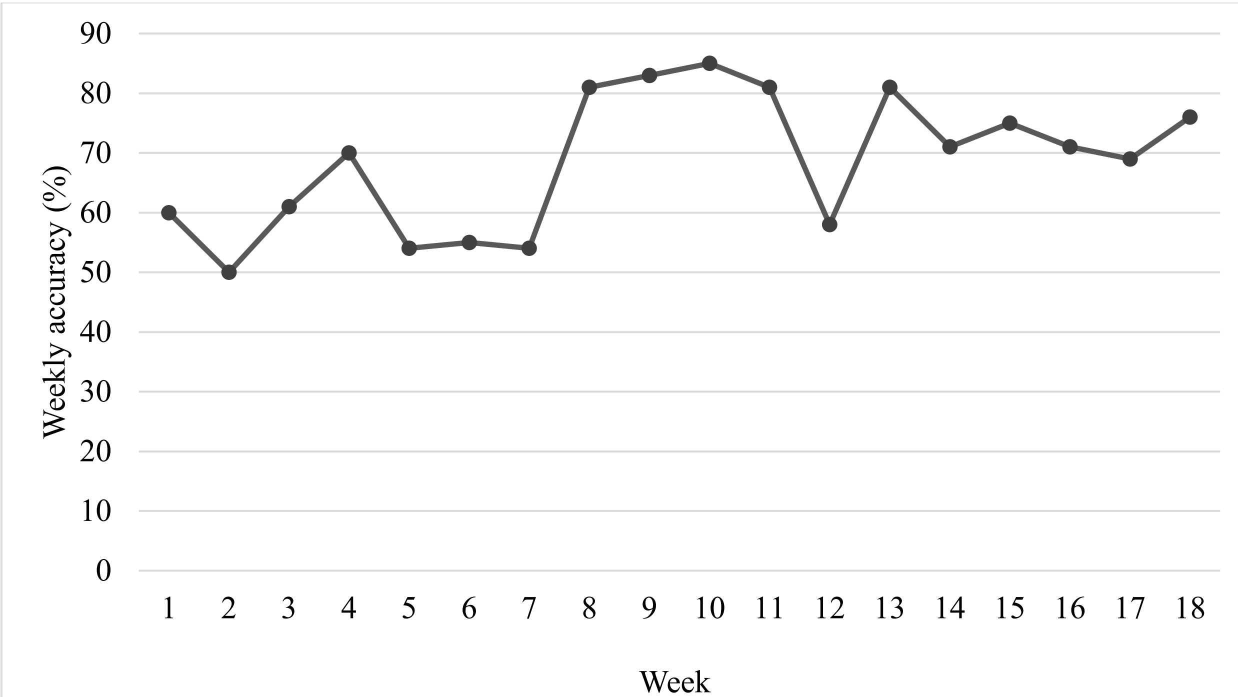
Week	Amount of data	Specificity		Precision		Sensitivity		*F*-Measure		Accuracy	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
1	113,585	99	100	95	100	6	6	11	11	22	16
2	443,040	55	63	87	91	60	60	71	72	59	60
3	726,397	70	68	91	89	61	58	73	70	62	59
4	973,496	71	77	90	93	58	57	71	71	60	61
5	1,203,686	69	69	91	92	70	69	79	79	69	69
6	1,572,754	79	85	92	95	59	62	72	75	63	66
7	1,872,702	83	85	93	94	53	51	68	66	59	57
8	2,336,105	69	71	92	92	83	78	87	84	80	77
9	2,492,457	81	79	93	93	63	57	75	71	66	61
10	2,614,689	56	55	89	90	89	88	89	89	82	82
11	2,776,942	77	72	93	92	74	72	82	81	75	72
12	3,028,478	79	78	93	93	72	67	81	78	74	69
13	3,255,458	59	65	89	91	89	84	89	87	83	80
14	3,441,195	71	73	92	93	84	79	88	85	82	78
15	3,667,310	74	78	92	94	80	74	86	83	79	74
16	3,992,200	76	76	93	94	78	77	85	85	78	77
17	4,241,995	73	72	92	94	82	82	87	88	80	80
18	4,468,906	61	60	90	91	90	89	90	90	84	84

Portability of the Prediction Model

We verified the prediction model accuracy against the learning data gathered from the students taking the Introduction to Artificial Intelligence distance course in the 2019–2020 academic year. The prediction model was assessed in terms of its accuracy in predicting the academic performance of the students in this new course; the results revealed an accuracy rate of ≥81% from the eighth week onward. The verification results of the prediction model are displayed in Figure 2.

Figure 2

Weekly Accuracy of the Verified Course



Discussion and Conclusions

Developing an EWS and identifying students at risk in a timely manner is one of the strategies of precision education for which schools and teachers have been advocating. Compared with face-to-face classes, distance courses enable the collection of more student learning information. However, for teachers who do not run MOOCs, gathering sufficient training information to build a usable prediction model themselves is a considerable challenge. The proportion of students who fail their SPOC is often higher than that of students who have face-to-face classes, especially for distance courses that use asynchronous teaching long term or during periods of special restrictions (e.g., contact restriction during a pandemic). Teachers' successful collection of small-sample learning information from multiple SPOCs and training of a portable prediction model would greatly benefit the development of an EWS, enabling teachers to employ precision education. This research is based on few-shot learning practice which feeds a predictive model with a very small amount of training data to discover patterns in data regarding accurate predictions. In this research, we gathered learning information from one teacher's multiple SPOCs on an LMS platform to create an EWS for identifying students at risk of failing. Our results revealed that students at risk can be correctly identified from the fifth week of the course onward on the basis of their online learning behavior (accuracy was 69%). The model's accuracy reached ≥ 80% for weeks 8, 10, 13, 14, 17, and 18. In this study, we obtained the accuracy of the confusion matrix to verify predictive models' performance. Additionally, the study also obtained the sensitivity, specificity, precision, and F measurement for each week to help teachers make comprehensive judgments when choosing different weekly patterns on the basis of their early warning plan. The main purpose of this research was to collect small-sample information from multiple SPOCs with a similar instructional design and taught by a single teacher to build a usable prediction model. Our findings help expand knowledge on the portability of prediction models and help confirm previous research that has indicated that the difference in instructional design between courses negatively affects the accuracy of student performance prediction. Therefore, teachers may use this prediction model in other distance courses that have similar online instructional designs and apply instructional interventions for students who are identified. Through instructional intervention, the online learning behavior of students taking SPOCs can be modified and their online learning experience enriched, such as through self-regulated learning. We endeavor to expand this research project by integrating automated data collection, feature selection, and model update mechanisms into the prediction model to enhance the model's adaptability and usability.

Practical Implications

In this study, we attempted to address a problem in EWS design: the necessity of first collecting big data on student learning before the development of early warning models. As a possible supporting technology, artificial intelligence has emerged in many industries. However, because of the lack of large data sets, educational institutions have yet to widely adopt this technology. In this context, teachers also miss the opportunity to develop predictive models for their SPOCs and cannot establish an EWS. Because teachers cannot directly supervise students' online learning behaviors as they would in the classroom, students who take online asynchronous courses are at an increased risk of failure.

The findings of this research may be of value to those who teach asynchronous distance courses, educational authorities, and information technology (IT) directors of academic institutions.

Teachers

Teachers should consider other factors in addition to online teaching design and regard the online learning environment as a sustainable and circular ecosystem. For example, in this study, we used former students' learning data sets and used a CNN to establish an early warning model to reduce future students' learning risk. This system is sustainable because new data can be integrated into the early warning model to improve its accuracy. In this manner, teachers can offer precision education through data-driven interventions. This system can support teachers in realizing the digital transformation of education. Such a system also enables teachers to devote more energy to supporting students' success in a timely and personalized manner.

Educational Authorities

Educational authorities should fine-tune their vision, draft policies, and provide funding for the development of learner-oriented artificial intelligence (AI) to enrich students' distance learning experiences and teacher effectiveness in SPOCs. For example, educational authorities could organize seminars to promote dialogue among university teachers, data analysts, and IT specialists. Administrators could also use case studies of successful AI applications in teaching as the basis for training materials to develop AI applications in distance education. Finally, relevant authorities could host conferences or workshops on the ethics of applying AI in education to enhance the knowledge of teachers and related personnel.

IT Directors

IT directors of academic institutions should establish systems that enable teachers to rapidly obtain LMS course data. For example, this could be done by establishing a learning data warehouse where online course data could be stored or providing an automatic access mechanism that gives teachers timely access to data (e.g., through an API). IT directors should also organize and publish descriptions of the data set, such as in a codebook.

Recommendations for Further Research

The data sets we used to build the EWS were all derived from a university in Taiwan. This research also preliminarily verified that the early warning model could be transferred to another course if its instructional design was similar to that of the source course. However, we did not further examine the uncertainty factors that may cause model migration to fail because of the bias in training data collection; this may arise for courses with multicultural learners or in the transfer of the model for use on students in other grades (e.g., K–12).

References

Baker, R. S. J. D., Costa, E., Amorim, L., Magalhães, J., & Marinho, T. (2012). Mineração de ados ducacionais: Conceitos, écnicas, erramentas e plicações (Educational data mining: Concepts, techniques, tools and applications. Updating day in informatics in education). *Jornada de Atualização em Informática na Educação*, *1*, 1–29.

Baker, R. S., Lindrum, D., Lindrum, M. J., & Perkowski, D. (2015). Analyzing early at-risk factors in higher education e-learning courses. In J. G. Boticario, O. C. Santos, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 150–155). http://www.columbia.edu/~rsb2162/2015paper41.pdf

Baradwaj, B. K., & Pal, S. (2011). Mining educational data to analyze students performance. *International Journal of Advanced Computer Science and Applications, 2*(6), 63–69. http://dx.doi.org/10.14569/IJACSA.2011.020609

Barry, M., & Reschly, A. (2012). Longitudinal predictors of high school completion. *School Psychology Quarterly, 27*(2), 74–84. https://doi.org/10.1037/a0029189

Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—A literature review. *SAGE Open, 6*(1), 1–11. https://doi.org/10.1177/2158244015621777

Bowles, M. (2015). *Machine learning in Python—Essential techniques for predictive analysis*. Wiley.

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

Chen, W., Brinton, C. G., Cao, D., Mason-Singh, A., Lu, C., & Chiang, M. (2018). Early detection prediction of learning outcomes in online short-courses via learning behaviors. *IEEE Transactions on Learning Technologies, 12*(1), 44–58. http://dx.doi.org/10.1109/TLT.2018.2793193

Conijn, R., Kleingeld, A., Matzat, U., Snijders, C., & van Zaanen, M. (2016, December 10). *Influence of course characteristics, student characteristics, and behavior in learning management systems on student performance* [Paper presentation]. Neural Information Processing Systems (NIPS) Workshop on Machine Learning for Education 2016, Barcelona, Spain.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17–29. https://doi.org/10.1109/TLT.2016.2616312

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior, 73*, 247–256. https://doi.org/10.1016/j.chb.2017.01.047

Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of Medical Systems, 43*(6), 162.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Goutte C., & Gaussier E. (2005). A probabilistic interpretation of precision, recall and *F*-score, with implication for evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Lecture Notes in Computer Science: Vol. 3408. Advances in Information Retrieval. ECIR 2005* (pp. 345-359). https://doi.org/10.1007/978-3-540-31865-1_25

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

He, H., & Garcia, E. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

He, J., Bailey, J., Rubinstein, B. I. P., & Zhang, R. (2015). Identifying at-risk students in massive open online courses. In B. Bonet & S. Koenig (Eds.), *Proceedings of the twenty-ninth AAAI conference on artificial intelligence* (pp. 1749–1755). AAAI. https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9696/9460

Howard, E., Meehan, M., & Parnell, A. (2018). Contrasting prediction methods for early warning systems at undergraduate level. *The Internet and Higher Education, 37*, 66–75. https://doi.org/10.1016/j.iheduc.2018.02.001

Hu, Y.-H., Lo, C.-L., & Shih, S.-P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469–478. https://doi.org/10.1016/j.chb.2014.04.002

Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience, 2018*, Article 6347186, 1–21. https://doi.org/10.1155/2018/6347186

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Jokhan, A., Sharma, B., & Singh, S. (2018). Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education, 44*(11), 1900–1911. https://doi.org/10.1080/03075079.2018.1466872

Kim, B., Vizitei, E., & Ganapathi, V. (2018). Gritnet: Student performance prediction with deep learning. In K. E. Boyer & M. Yudelson, (Eds.), *Proceedings of the 11th international conference on educational data mining*. International Educational Data Mining Society. https://arxiv.org/abs/1804.07405

Kotsiantis, S. B. (2011). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review, 37*(4), 331–344. https://doi.org/10.1007/s10462-011-9234-x

Kotsiantis, S. B., Pierrakeas, C., & Pintelas, P. (2004). Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, *18*(5), 411–426. https://doi.org/10.1080/08839510490442058

Lauría, E. J. M., Baron, J. D., Devireddy, M., Sundararaju, V., & Jayaprakash, S. M. (2012). Mining academic data to improve college student retention. In S. Dawson, C. Haythornthwaite, S. B. Shum, D. Gašević, & R. Ferguson (Eds.), *LAK '12: Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 139–142). ACM. https://doi.org/10.1145/2330601.2330637

Li, X., Wang, T., & Wang, H. (2017). Exploring N-gram features in clickstream data for MOOC learning achievement prediction. In Z. Bao, G. Trajcevski, L. Chang, & W. Hua (Eds.), *Lecture Notes in Computer Science: Vol. 10179. Database Systems for Advanced Applications. DASFAA 2017* (pp. 328–339). https://doi.org/10.1007/978-3-319-55705-2_26

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Milne, J., Jeffrey, L. M., Suddaby, G., & Higgins, A. (2012). Early identification of students at risk of failing. In M. Brown, M. Harnett, & T. Stewart (Eds.), *Future challenges, sustainable futures: Proceedings ASCILITE* (pp. 657–661). http://www.ascilite.org/conferences/Wellington12/2012/images/custom/milne,_john_-_early_identification.pdf

Morris, L. V., Finnegan, C., & Wu, S.-S. (2005). Tracking student behavior persistence and achievement in online courses. *The Internet and Higher Education, 8*(3), 221-231. https://doi.org/10.1016/j.iheduc.2005.06.009

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. In A. Wise, P. H. Winne, G. Lynch, X. Ochoa, I. Molenaar, S. Dawson, &

M. Hatala (Eds.), *LAK '17: Proceedings of the 7th international learning analytics and knowledge conference* (pp. 598–599). ACM. https://doi.org/10.1145/3027385.3029479

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society, 17*(4), 49–64. https://www.jstor.org/stable/pdf/jeductechsoci.17.4.49

Ramaswami, M., & Bhaskaran, R.(2009). A study on feature selection techniques in educational data mining. *Journal of Computing, 1*(1), 7–11. https://arxiv.org/abs/0912.3924

Rienties, B., Toetenel, L., & Bryan, A. (2015). 'Scaling up' learning design: Impact of learning design activities on LMS behavior and performance. In J. Baron, G. Lynch, N. Maziarz, P. Blikstein, A. Merceron, & G. Siemens (Eds.), *LAK '15: Proceedings of the 5th international conference on learning analytics and knowledge* (pp. 315–319). https://doi.org/10.1145/2723576.2723600

Rodrigues, R. L., Ramos, J. L. C., Sedraz, J., & Gomes, A. S. (2016). Discovery engagement patterns MOOCs through cluster analysis. *IEEE Latin America Transactions, 14*(9), 4129–4135. https://doi.org/10.1109/TLA.2016.7785943

Romero, C., & Ventura, S. (2013). Data mining in education. *WIREs, 3*(1), 12–27. https://doi.org/10.1002/widm.1075

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

Saa, A. A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. *Technology, Knowledge and Learning, 24*, 567–598.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE, 10*(3), 1–21. https://doi.org/10.1371/journal.pone.0118432

Sana, B., Siddiqui, I. F., & Arain, Q. A. (2019). Analyzing students' academic performance through educational data mining. *3C Tecnología. Glosas de innovación aplicadas a la pyme, 29*(2), 402–421. http://dx.doi.org/10.17993/3ctecno.2019.specialissue2.402-421

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157–167. https://doi.org/10.1016/j.chb.2014.05.038

Toraman, S., Tuncer, S. A., & Balgetir, F.(2019). Is it possible to detect cerebral dominance via EEG signals by using deep learning? *Medical Hypotheses, 131*, Article 109315. https://doi.org/10.1016/j.mehy.2019.109315

Wolff, A., Zdrahal, Z., Herrmannova, D., Kuzilek, J., & Hlosta, M. (2014). Developing predictive models for early detection of at-risk students on distance learning modules. In M. Pistilli, J. Willis, D. Koch, K. Arnold, S. Teasley, & A. Pardo (Eds.), *LAK '14: Proceedings of 4th international learning analytics and knowledge conference* (pp. 24–28). http://oro.open.ac.uk/40669/

International Review of Research in Open and Distributed Learning
Volume 23, Number 1

February – 2022

MOOC Evaluation System Based on Deep Learning

Jian-Wei Tzeng[1], Chia-An Lee[2], Nen-Fu Huang[2], Hao-Hsuan Huang[2], and Chin-Feng Lai[3*]
[1]Center for Teaching and Learning Development, National Tsing Hua University, Taiwan; [2]Department of Computer Science, National Tsing Hua University; [3]Department of Engineering Science, National Cheng Kung University, Taiwan
[*]The corresponding author is Chin-Feng Lai

Abstract

Massive open online courses (MOOCs) are open access, Web-based courses that enroll thousands of students. MOOCs deliver content through recorded video lectures, online readings, assessments, and both student–student and student–instructor interactions. Course designers have attempted to evaluate the experiences of MOOC participants, though due to large class sizes, have had difficulty tracking and analyzing the online actions and interactions of students. Within the broader context of the discourse surrounding big data, educational providers are increasingly collecting, analyzing, and utilizing student information. Additionally, big data and artificial intelligence (AI) technology have been applied to better understand students' learning processes. Questionnaire response rates are also too low for MOOCs to be credibly evaluated. This study explored the use of deep learning techniques to assess MOOC student experiences. We analyzed students' learning behavior and constructed a deep learning model that predicted student course satisfaction scores. The results indicated that this approach yielded reliable predictions. In conclusion, our system can accurately predict student satisfaction even when questionnaire response rates are low. Accordingly, teachers could use this system to better understand student satisfaction both during and after the course.

Keywords: MOOC, deep learning, learner satisfaction, learning analytics

Introduction

Massive open online courses (MOOCs) are open-access educational resources that offer various academic courses to the general public through the Internet (Kop, 2011). Since 2012, MOOCs have included high-quality video lectures from universities worldwide. The self-directed learning environment provided by MOOCs signifies a modern approach to education. Users of MOOCs can learn not only from instructional videos created by professors but also through other methods suited to their individual learning styles, including live-streaming video lectures, efficient assessments, and discussion forums (McAuley et al., 2010).

A considerable amount of learning data can be collected and analyzed from the increasingly large number of MOOC users. Many studies have been conducted based on MOOC data; for instance, Kop et al. (2011) described the use of Facebook groups by MOOC participants and obtained data from learner experience surveys, participant demographics, and learner progression through courses. Adamopoulos (2013) analyzed a dataset of MOOC user-generated content to identify factors that predicted self-reported course progress.

Within the broader context of the discourse surrounding big data, educational providers are increasingly collecting, analyzing, and using student information (Papamitsiou & Economides, 2014; Su et al., 2021; Su & Lai, 2021; Su & Wu, 2021). Data have been collected to personalize learning experiences and allocate resources to individual students (Gašević et al., 2015; Leitner et al., 2017). Additionally, big data and artificial intelligence (AI) technology have been applied to better understand learning. Researchers initially focused on creating personalized teaching systems for lone learners, but recent studies have emphasized the interactions between students and the learning material (Kay, 2012). Cognitive science can be used to help lecturers understand the nature of learning and teaching. Thus, the findings can be used to build better systems to help learners gain new skills or understand new concepts. AI has now begun to affect the student experience through analyses of learning data (du Boulay, 2016).

Learner satisfaction refers to student perceptions of both the learning experience and the value of the education received (Baxter Magolda, 1993). According to Donohue and Wong (1997), satisfaction can affect student motivation. It is a significant intermediate outcome (Donohue & Wong, 1997) and a predictor of retention (Baxter Magolda, 1993). Bean and Bradley (1986) found that for college students, satisfaction had a greater impact on their performance than performance had on their satisfaction. However, Klobas et al. (2014) stated that researchers know very little about learner motivations, experiences, and satisfaction. Veletsianos (2013) also noted that discussions about new educational innovations, such as MOOCs, lack input from learners. Accordingly, it is reasonable to conclude that student satisfaction, as determined by student feedback, is a critical factor influencing academic success.

Some studies, such as Liu et al. (2014) and Onah et al. (2014) have characterized MOOC student perspectives by investigating what they learned, the aspects of MOOCs they found most useful, and their motivations for enrolling in MOOCs. However, these studies have been limited to surveying enrollees in journalism MOOCs or analyzing blog posts written by MOOC students related to their MOOC experiences.

Researchers have tried to understand the high dropout rate of MOOCs. (Magold, 1993). Onah et al. (2014) postulated several reasons of low dropout rate, such as low motivation to complete the courses, lack of time, digital and learning skills, and level of the course and lack of support.

Information collected by researchers and e-learning providers has come primarily in the form of big data or learning analytics gathered from observations of online student interactions with the instructors, the content, and their classmates. However, this approach has proved insufficient for gaining a comprehensive understanding of learner experiences in open online learning.

Studies investigating MOOCs from the perspective of an individual learner have collected data from learner experience surveys and on (a) participant demographics; (b) learner progression throughout various courses (in terms of, for example, the number of videos viewed or tests taken; Kop et al., 2011); (c) class size and completion rate (Adamopoulos, 2013); or (d) students' behaviors, motivations, and communication patterns (Swinnerton et al., 2016). These metrics mirrored attendance and completion data and have enabled researchers to assess this form of education.

Advancements in technology have enabled the application of data-mining techniques and AI to the analysis of MOOCs. Some studies of MOOC performance have analyzed the language used in discussion forums to make predictions. Other researchers have used natural language processing (McNamara et al., 2015; Wen et al., 2014). More recently, these techniques have been used to identify student sentiment among MOOC enrollees (Moreno-Marcos et al., 2018; Pérez et al., 2019).

Due to its numerous advantages, AI has been increasingly applied in education. First, AI techniques have improved lecturers' understanding of learning and teaching, and facilitated the design of new systems that help learners gain new skills or grasp new concepts (du Boulay, 2016). Therefore, the application of AI to large MOOC datasets has drawn substantial attention. Second, Fauvel et al. (2018) proposed that AI tools could be used to better understand MOOC participant sentiment, and that MOOC instructors use these data to deliver better courses and develop more useful educational tools. AI could also be used to analyze student learning effectiveness by using records of learning behaviors. Some AI tools have been applied to make online learning more similar to its offline counterpart in order to help students better achieve their learning goals. Because of the variety in student learning adaptability, habits, and behavior, personalized service in MOOCs has been seen as especially important (Tekin et al., 2015).

Although there has been an increasing interest in artificial intelligence in educational research, less than five percent of such studies have addressed deep learning in education. However, given the rapid advance of deep learning, application of it in education is seen to have dramatic potential (Chen et al., 2020). Therefore, in terms of future research, the system examined in this study, since it is based on deep learning, could be a useful example of developing such a system for predicting student performance.

One of the challenges of lecturing in a MOOC is accurately understanding the learner experience. It has proved impossible to keep track of all posts and interactions of the numerous enrollees. The analysis of individual learner experience is critical for course evaluation. According to Donath (1996) learner comments and actions indicated their sentiments and concerns toward a course. Without the appropriate analytical tools, it has been difficult to understand differences in learner sentiment and experiences across different learner groups in a large class.

This study proposed a method for evaluating students' satisfaction by using machine learning. In this method, the learning behavior of participants within the course was used as input for the model, and compared with the results of a survey of MOOC students. The method focused specifically on certain

MOOC features students considered important. Thus, educators can use the findings of this research in order to modify their MOOCs to increase student satisfaction and enhance the student learning experience.

Training data for the model came from MOOCs at National Tsing Hua University (NTHU). Logs of learning activities, such as video-watching behavior and exercise completion, were collected and transformed to measures of learning behavior in the model. The proposed model used a deep neural network (DNN) with regression. The result predicted by the DNN was compared with survey responses to evaluate the accuracy of the model. These findings helped us evaluate MOOC learner satisfaction, and aided the design and execution of MOOC lectures.

Student Feedback

Student feedback to the courses is one of significant indicators in both face-to-face and online courses. Due to the availability of educational big data, Gameel (2017) analyzed data collected from 1,786 learners enrolled in four MOOCs. Learners perceived that the following aspects influenced learning satisfaction: learner–content interaction, as well as the usefulness, teaching aspects, and learning aspects of the MOOC. From learners' perspective, those aspects offer valuable insights into understanding the quality and satisfaction of the MOOCs.

To date, MOOCs have not provided participants (i.e., educators or learners) with any form of timely analysis on forum content. Consequently, educators have been unable to reply to questions or comments from hundreds of students in a timely manner (Shatnawi et al., 2014).

Because feedback has been too general, incomplete, or even incorrect, automation may be a solution to this problem. Automatic techniques include (a) functional testing, where feedback is usually insufficient as a guide for novices; (b) software verification for finding bugs in code, which may confuse novices because these tools often ignore true errors or report false errors; and (c) comparisons using reference solutions, in which many reference solutions or pre-existing correct submissions are usually required. One study used a semantic-aware technique to provide personalized feedback that aimed to mimic an instructor looking for code snippets in student submissions for a coding MOOC (Marin et al., 2017).

Moreover, some researchers take advantage of machine learning to analyze the feedback from MOOCs (Hew et al., 2020). Several deep learning models are used to predict student performance, such as dropout prediction (Xing & Du, 2019) or grade prediction (Yang et al., 2017). To make the accuracy higher, precise big data analysis is also a critical direction thing to MOOC. Some researchers want to analyze video watching data precisely (Hu et al., 2020).

Higher education institutions and experts have had a strong interest in extracting useful features pertaining to the course and to learner sentiment from such feedback (Dohaiha et al., 2018; Kastrati et al., 2020). It is thus imperative to develop a reliable automated method to extract these sentiments when dealing with large MOOCs (Sindhu et al., 2019). For instance, Lundqvist et al. (2020) evaluated student feedback within a large MOOC. Their dataset contained 25,000 reviews from MOOC users. The participants were divided into three groups (i.e., beginner, experienced, and unknown) based on their level of experience with the topic. The researchers used the Valence Aware Dictionary for Sentiment Reasoning as an algorithm for sentiment analysis.

Course Evaluation

Several studies were instructive sources for the design of the questionnaire used in this research. Durksen et al. (2016) used cutting-edge methods to analyze students' satisfaction in a learning environment. They examined educational and psychological aspects of traditional and MOOC learning settings to compare outcomes (e.g., students' characteristics, course design). This psychological perspective postulated that the basic needs for autonomy, competence, relatedness, and belonging characterized learner experiences in MOOCs (Durksen et al., 2016).

Other studies have focused on workload and precisely quantified students' workload. In one study, the workload of medical students was quantized using a specifically developed and self-completed questionnaire (Gonçalves, 2014). Additionally, Çakmak (2011) designed a method to quantify instructor style, including factors such as making clear statements, using one's time effectively, and using technology. Çakmak referred to student positivity towards instructor style as style approval. Marciniak (2018) also described effective methods for assessing course quality, which encompassed dimensions evaluating all aspects of the program.

Research Design

Below, we first describe the data collection process in terms of course information and learning behavioral data used in this study. Examples of schema of video and exercise from the platform are also shown to indicate the data structure. Then, we report the design and content of student questionnaire with the response rate of each course. Finally, how data is extracted from the learning activity logs to formulate the predictive model is illustrated with performance evaluation measure.

Course Information

To avoid bias, different types of MOOC courses offered by NTHU in February 2020 were selected:

- *Introduction to IoT (Internet of Things)*
- *Introduction to Calculus*
- *Introduction to Programming in Python*
- *Financial Decision Analysis*
- *Systems Neuroscience*
- *Ecosystem and Global Changes*
- *Common Good in Social Design*
- *Introduction to Data Structure*

Some advanced placement (AP) courses offered in May 2020 were also chosen:

- *AP-Introduction to Calculus*

- *AP-General Physics*
- *AP-General Chemistry*
- *AP-Introduction to Life Science*
- *AP-Principles of Economics*
- *AP-Introduction to Computer Science*
- *AP- Introduction to Programming in Python*
- *AP-Introduction to Computer Programming*

Students in these MOOCs were expected to spend three hours each week watching online videos and completing practice exercises. They were also expected to discuss the course content with their peers. For *Introduction to IoT*, students were also required to conduct experiments in some offline laboratory sessions.

Collection of Learning Behavior Data

Videos are the primary teaching method for most MOOCs. In this study, we collected data on video playback actions, such as playing, pausing, seeking, and adjusting the playback speed (Table 1) as well as data on each user's answers for each exercise (Table 2). If a student entered the exercise page but did not answer the exercise questions, we coded the student's response to the exercise as *No*. The feature timeCost (in seconds) was defined as how long the student took to answer each question. For example, if a student spent 10 seconds answering a question, the timeCost value for that student for that question was 10. The 308,517,712 learning behavior data was transformed into meaningful features as input of the DNN model. We sorted all course data into the categories of training data, validation data, and testing data according to the ratio of 0.64, 0.16, and 0.2.

Table 1

Student Video Activity Schema

Feature	Description	Example
userId	Student ID	2,198
courseId	Course ID	10900CS0003
chapterId	Chapter ID	10900CS0003ch04
videoId	Video ID	-WSFgrGEs
action	Student action when recording	playing
currentTime	Video time when recording	13.3234
playRate	Video play rate when recording	1.25
volume	Video volume when recording	100
update_at	Recording time	2020-05-11T22:40:41

Table 2

Student Exercise Activity Schema

Feature	Description	Example
userId	Student ID	2198
courseId	Course ID	10900CS0003
chapterId	Chapter ID	10900CS0003ch04
exerId	Exercise ID	10900CS0003ch04e1
score	Exercise answer score	1
timeCost	Time cost on this exercise	10
userAns	Student answer	[1, 3, 4]
correctAns	Correct answer	[1, 2]
update_at	Recording time	2020-05-11T23:51:41

Survey Questionnaire

Referencing the literature, we focused on the following five categories of student sentiment survey: (a) workload, (b) need fulfillment, (c) intelligibility, (d) style approval, and (e) student engagement. The questionnaire had 22 items in total (Table 3). This research used five-point Likert scale to evaluate the answers provided by students. Rating 1 indicated their worst experience while rating 5 indicated their best experience.

Of the 6,016 students enrolled in the aforementioned courses, 993 filled out the questionnaire, and 764 of the 993 responses were valid. The questionnaire response rates for each course are reported in Table 4; the Cronbach's alpha was 0.842. The response rates for the various courses ranged from 5% to 15%, a result strongly correlated with the number of students completing their MOOCs (Jordan, 2015). *Introduction to IoT* had the highest response rate (45%) due to the requirement for learners to attend in-person experiment sessions.

Table 3

Questionnaire Design and Content

Field	Citation	Topic
Workload	Gonçalves (2014)	It takes a lot of time to watch the videos for this course.
		I think this course is quite difficult.
		I can keep up with the subsequent courses without spending much time reviewing.
Need fulfillment	Durksen et al. (2016)	The course material is consistent with what I expect to learn.
		The course material is not what I currently need to learn.

	Alario-Hoyos et al. (2017)	This course will be helpful for my future courses and research.
		This course is helpful for my future job search.
		This course is related to my major.
Intelligibility	Ochando (2017)	The teacher's style helps me easily understand the content.
		The teacher is able to explain the key points and clarify confusing points.
		The teacher's method is too disorganized for me to keep up.
		The teacher is unclear, and I have difficulty understanding.
		The teacher's methods make me feel that this course is an efficient way to learn.
Style approval	Çakmak (2011)	The teacher's style makes me eager to learn.
		The way the teacher speaks makes me feel a little hesitant.
		The teacher's tone does not make me feel irritated.
		The teacher's rhythm puts me at ease.
		The teacher's methods make me feel pressured.
Student engagement	Marciniak (2018)	I watched the course videos at least once before the end of the course.
		I review the exercises by myself offline.
		I will find related videos about unfamiliar concepts.
		I will rewatch videos to review unfamiliar concepts.

Table 4

Response Rate Information

Course name	Number of students	Number of responses	Response rate
Introduction to IoT	255	115	0.45
Introduction to Calculus	490	103	0.20
Introduction to Programming in Python	569	95	0.16
Financial Decision Analysis	383	56	0.14
Systems Neuroscience	201	22	0.10
Ecosystem and Global Change	233	43	0.18

Common Good in Social Design	121	20	0.16
Introduction to Data Structure	249	26	0.10
AP-Introduction to Calculus	980	217	0.22
AP-General Physics	275	26	0.09
AP-General Chemistry	371	52	0.14
AP-Introduction to Life Science	156	27	0.17
AP-Principles of Economics	172	17	0.10
AP-Introduction to Computer Science	202	18	0.08
AP-Introduction to Programming in Python	449	58	0.12
AP-Introduction to Programming Language	259	22	0.08

Extracting Data on Features Related to Learning Activities

Logs of activity involving videos and exercises were collected. Information regarding video playback fell into one of seven categories: (a) video operations (e.g., play and pause); (b) whether the users actually watched the video being played; (c) start and end times; (d) current time; (e) playback speed; (f) volume; and (g) other information. We analyzed the information and extracted 35 video features, as listed in Table 5. Information regarding exercises was divided into six categories: (a) user answer, (b) correct answer, (c) points, (d) time spent by users, (e) types of questions, and (f) other. Eight exercise features were extracted (Table 5).

Table 5

Video and Exercise Features

Feature	Description
real_watch_time	Time spent watching videos (including time spent with the video on pause or fast-forwarded)
video_watch_time	Time spent watching videos
play_count	Number of times the video was played in a week
pause_count	Number of times the video was paused in a week
change_rate_count	Number of times the video play rate was changed in a week
seek_forward_count	Number of times the video was fast-forwarded in a week
seek_back_count	Number of times the video was skipped backward in a week
finish_ratio	Ratio of videos finished in a week
review_ratio	Ratio of videos reviewed in a week

video_progress_ratio	Proportion of video play time and video watch time
video_len_per_week	Total length of videos assigned per week
real_watch_time_per_week	Total video watch time per week (included pause, fast-forward, and others)
video_watch_time_per_week	Video watch time per week
real_watch_time_per_video_len	Proportion of student's watch time to total watch time of all assigned videos in a week (watch time included time spent with the video on pause or fast-forwarded)
video_watch_time_per_video_len	Proportion of all assigned videos that were watched in a week
end_to_start_days	Days from learning start date to learning end date
real_learning_days	Numbers of learning days per week
times_in_real_learning_days	Time spent learning during a learning day
average_learning_time_each_path	Average learning time for a student
week_block_num_mean	Mean number of learning blocks in a week
week_block_num_std	Standard deviation of mean number of learning blocks in a week
week_block_time_mean	Mean time of learning blocks in a week
week_block_time_std	Standard deviation of number of learning blocks in a week
day_block_time_mean	Mean number of learning blocks in a learning day
day_block_num_std	Standard deviation of mean number of learning blocks in a learning day
day_block_time_mean	Mean time of learning blocks in a learning day
day_block_time_std	Standard deviation of number of learning blocks in a learning day
min_15	Mean number of learning blocks > 15 min. in a week
min_30	Mean number of learning blocks > 30 min. in a week
min_45	Mean number of learning blocks > 45 min. in a week
gap_mean	Mean number of days spent not learning
gap_std	Standard deviation of the number of days spent not learning
course_weeks	Number of times a student took a course in a week

video_total_len	Sum of the lengths (in seconds) of all videos in a week
video_counts	Number of videos watched in a week
exercise_type	Exercise type (single, multiple, or fill-in-the-blanks)
correct_rate	Percentage of questions answered correctly
answer_count	Number of attempts before the student answered correctly
time_cost	Time taken to complete an exercise
review_video_before_answer	Whether the student watched a related video before answering correctly for the first time
review_video_after_answer	Whether the student watched a related video after answering correctly for the first time
answering_progress	Type of question processing style (type 1 to 6)
correct_count	Number of questions answered correctly

Prediction of Questionnaire Score Based on Learning Behaviors

Every student has a unique learning mode and unique learning behavior, and we hypothesized that these would affect their satisfaction. To verify this hypothesis, we inputted the student learning behavior variables into a five-layer DNN model (see Figure 1 for illustration), which used a rectified linear unit activation function to predict the satisfaction score. When creating predictions of student satisfaction for MOOCs, it is crucial to avoid inaccuracies caused by sparse data (Yang et al., 2018). To avoid this problem, input for our system included only the learning data of students who completed the questionnaire.

Performance Evaluation for the System

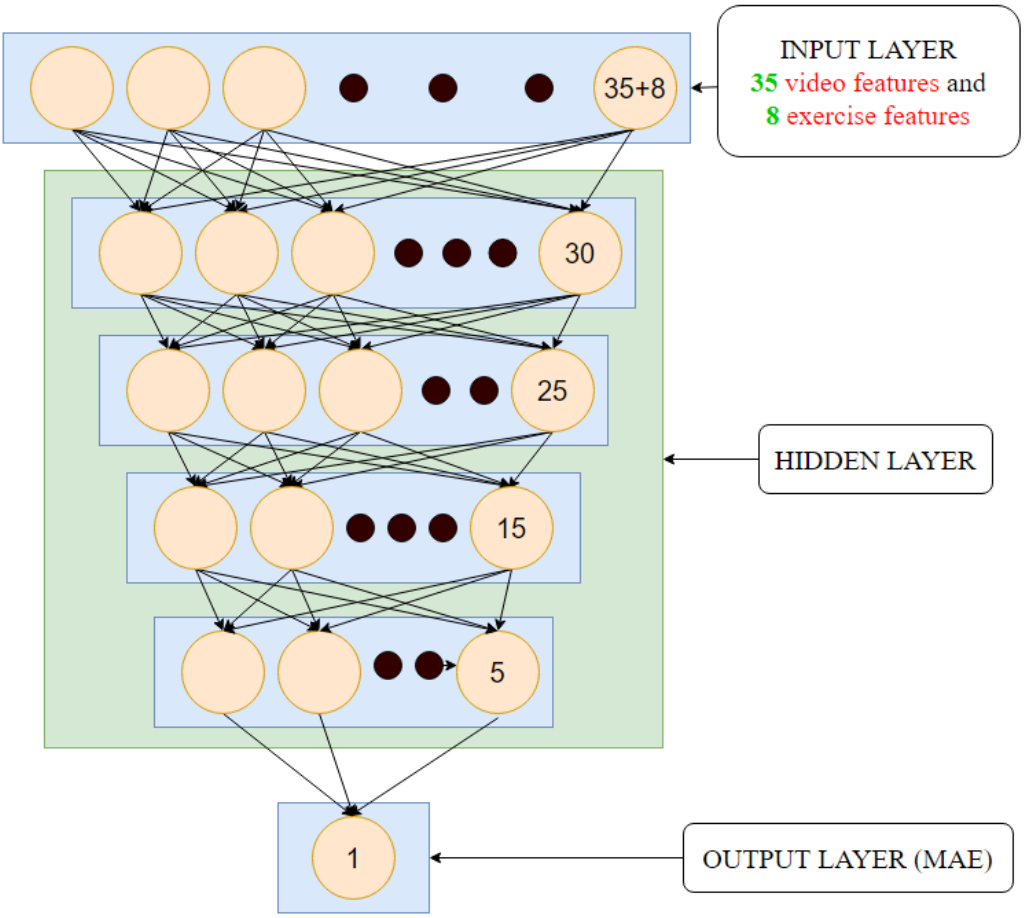
The mean absolute error (MAE) was used to evaluate the performance of the model. In brief, we used holdout cross validation to obtain the test data, and the data were then used to calculate the MAE as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - y_i|, \tag{1}$$

where f_i and y_i are the predicted and actual scores for student i, respectively, and N is the number of students. The MAE is the difference between the predicted and actual scores, with a lower MAE indicating superior predictive performance.

Figure 1

Architecture of Satisfaction Score Prediction Model Based on Learning Behaviors

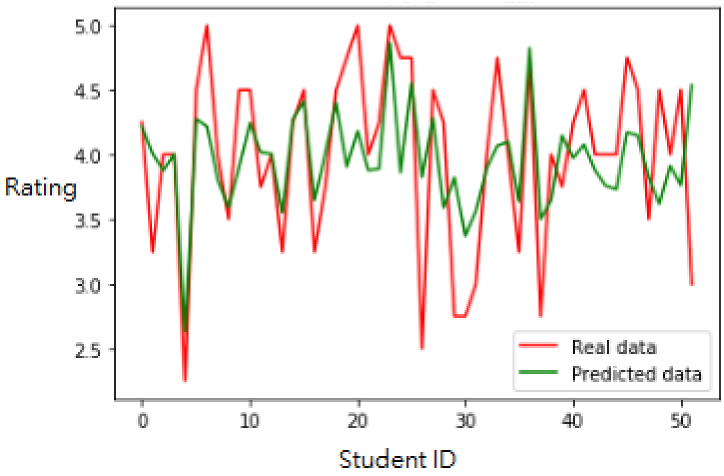


Results and Discussion

The effectiveness of our prediction model was evaluated in terms of the MAE by using the data from the Table 4 courses. Table 6 shows the MAE output for the answer to each question from the questionnaire. Our model performed best when computing the answers to questions related to course health, and worst when computing the answers to style approval questions. The results indicated that learning behavior is affected to some degree by student satisfaction. The MAEs shown in Table 6 ranged from 0.41 to 0.55. This may result from our use of a five-point rating system. The predicted data successfully captured the trend of real data, as shown in Figure 2.

Figure 2

Example of Satisfaction Score Prediction Model Based on Learning Behaviors



Note. This figure is made based on Question 5-1.

Once students' answers to the questionnaire survey were collected, the five categories of results were computed by the overall answer score for each part of the questionnaire. Subsequently, we collected data on the learning behavior of participating MOOC students. Thereafter, these data were analyzed and used to predict the student satisfaction scores.

The results demonstrated that this system enabled teachers to understand multiple aspects of learner satisfaction before the end of the course. Additionally, because course evaluation surveys have high nonresponse rates (Table 4) this system was useful as an alternative method of providing lecturers with feedback predictions for students who do not fill out questionnaires. On the basis of the predicted feedback, teachers can adjust the content, workload, teacher-student or student-student interactions during the course. Compared with the conventional approach, which is disadvantaged by insufficient learner responses and where feedback is given only after the course, our method was more flexible and accurate.

Before the end of the course, the instructor can also use different approaches to track student performance and thus help students by adjusting the course schedule, offering more office hours, or allocating more time to covering more difficult topics. In addition, this system may provide students a chance to reflect on their own performance based on the predictions.

In the future, this system could be combined with a learning log feature. Teachers would then use the student's learning history to better understand their status, and so develop more sophisticated and efficient interactive teaching methods, improve course quality, and increase student satisfaction.

Table 6

Predictive Performance

Field	Question (value of answers range from 1 to 5)	MAE
Workload	It takes a lot of time to watch the videos for this course.	0.4679
	I think this course is quite difficult.	0.5192
	I can keep up with subsequent courses without spending much time reviewing.	0.4679
Need fulfillment	The course material is consistent with what I expect to learn.	0.4308
	The course material is not what I currently need to learn.	0.4692
	This course will be helpful for my future courses and research.	0.4115
	This course is helpful for my future job search.	0.4654
	This course is related to my major.	0.4462
Intelligibility	The teacher's style helps me easily understand the content.	0.4654
	The teacher is able to explain the key points and clarify confusing points.	0.4691
	The teacher's method is too disorganized for me to keep up.	0.4769
	The teacher is unclear and I have difficulty understanding.	0.5231
	The teacher's methods make me feel that this course is an efficient way to learn.	0.5077
Style approval	The teacher's style makes me eager to learn.	0.4577
	The way the teacher speaks makes me feel a little hesitant.	0.5423
	The teacher's tone does not make me feel irritated.	0.523
	The teacher's rhythm puts me at ease.	0.5385
	The teacher's methods make me feel pressured.	0.5
Student engagement	I watched the course videos at least once before the end of the course.	0.4712
	I review the exercises by myself offline.	0.4135
	I will find related videos about unfamiliar concepts.	0.4136
	I will re-watch videos to review unfamiliar concepts.	0.4615

Limitations

The data used as input was collected from the courses in Table 4. Differences between these courses may affect the accuracy of our model. Future research might divide courses into categories to investigate subject matter–related effects. For example, the difficulty of a course may influence student concentration. Researchers can also use different methods to analyze the survey responses.

Conclusion

Education is foundational to a well-functioning society. Due to recent technological advancements, techniques from big data are now available for increasing the quality of courses. To properly use big data, researchers have adopted AI to investigate topics related to education. Through data analysis, processing, and prediction, AI can support lecturers in solving problems and making decisions. In combination with MOOCs, AI can help teachers create a better learning environment and enable students to achieve their learning goals—the common aim of all mainstream MOOC platforms.

In this study, we proposed a method to solve the problem of low MOOC student survey response rates, which prevents teachers from evaluating learner satisfaction in their courses. We established a system that predicted student course satisfaction based on their learning behavior. Our system was tested with student data from NTHU's MOOC platform. These data pertained to student behavior when watching videos and answering exercise questions. Subsequently, a deep learning model was used to process the data and produce a predicted level of course satisfaction for a given student. If the output is made viewable by students, this system may also give them a chance to reflect on their course performance based on the system's predictions.

Lastly, this system can benefit both lecturers and learners. Teachers can track student course satisfaction and learners can give instant feedback on course modifications. If a lecturer receives prompt feedback that guides course modifications, lecturers can better react to student input. Therefore, our system is an innovative method for improving interaction between teachers and students.

Acknowledgements

This study was funded by the Ministry of Science and Technology of Taiwan. The projects ID were MOST 110-2511-H-007 -007, MOST-109-2511-H-007-010 and MOST-108-2221-E-007-062-MY3. It was also funded by National Tsing Hua University, Taiwan.

References

Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. In R. Baskerville & M. Chau (Eds.), *Proceedings of the 34th International Conference on Information Systems (ICIS 2013)*. http://aisel.aisnet.org/icis2013/proceedings/BreakthroughIdeas/13/

Alario-Hoyos, C., Estévez-Ayres, I., Pérez-Sanagustín, M., Delgado Kloos, C., & Fernández-Panadero, C. (2017). Understanding learners' motivation and learning strategies in MOOCs. *The International Review of Research in Open and Distributed Learning, 18*(3), 120–-137. https://doi.org/10.19173/irrodl.v18i3.2996

Baxter Magolda, M. B. (1993). What *doesn't* matter in college? [Review of the book *What matters in college: Four critical years revisited*, by A. W. Astin]. *Journal of Higher Education 22*(8), 32–34. https://doi.org/10.2307/1176821

Bean, J., & Bradley, R. (1986). Untangling the satisfaction-performance relationship for college students. *The Journal of Higher Education, 57*(4), 393–-412. https://doi.org/10.2307/1980994

Boulay, B. d. (2016). Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems, 31*(6), 76-81. http://users.sussex.ac.uk/~bend/papers/meta-reviewsIEEEV5.pdf

Cakmak, M. (2011). Changing roles of teachers: Prospective teachers' thoughts. *Egitim ve Bilim, 36*, 14–-24. https://www.webofscience.com/wos/woscc/full-record/WOS:000286914500002?SID=E3dPJYYg8zalgwNi8L7 http://egitimvebilim.ted.org.tr/index.php/EB/article/download/173/235

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Dohaiha, H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2018). Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications, 118*, 272–-299. https://doi.org/10.1016/j.eswa.2018.10.003

Donath, J. (1996). Identity and deception in the virtual community. In M. H. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 29–59). Routledge. https://smg.media.mit.edu/people/judith/Identity/IdentityDeception.html

Donohue, T. L., & Wong, E. H. (1997). Achievement motivation and college satisfaction in traditional and nontraditional students. *Education, 118*(2), 237–-243. https://go.gale.com/ps/anonymous?id=GALE%7CA20479498&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00131172&p=AONE&sw=w

Du Boulay, B. (2016). Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems, 31*, 76–-81. https://doi.org/10.1109/MIS.2016.93https://www.computer.org/csdl/magazine/ex/2016/06/mex2016060076/13rRUxlgxVq

Durksen, T., Chu, M.-W., Ahmad, Z. F., Radil, A., & Daniels, L. (2016). Motivation in a MOOC: A probabilistic analysis of online learners' basic psychological needs. *Social Psychology of Education, 19*, 241–260. https://doi.org/10.1007/s11218-015-9331-9

Fauvel, S., Yu, H., Miao, C., Cui, L., Song, H., Zhang, L., Li, X., & Leung, C. (2018). *Artificial Intelligence Powered MOOCs: A Brief Survey* [Paper presentation]. IEEE International Conference on Agents. Singapore. https://doi.org/10.1109/AGENTS.2018.8460059https://ieeexplore.ieee.org/document/8460059

Gameel, B. (2017). Learner satisfaction with massive open online courses. *American Journal of Distance Education, 31*(2), 98–111. https://doi.org/10.1080/08923647.2017.1300462

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gonçalves, G. (2014). Studying public health. Results from a questionnaire to estimate medical students' workload. *Procedia - Social and Behavioral Sciences, 116*, 2915–2919. https://doi.org/10.1016/j.sbspro.2014.01.679

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education, 145*, 103724. https://doi.org/10.1016/j.compedu.2019.103724

Hu, H., Zhang, G., Gao, W., & Wang, M. (2020). Big data analytics for MOOC video watching behavior based on Spark. *Neural Computing and Applications, 32*(11), 6481-6489. https://doi.org/10.1007/s00521-018-03983-z

Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning, 16*(3). https://doi.org/10/f5vw

Kastrati, Z., Imran, A., & Kurti, A. (2020). Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. *IEEE Access, 8*, 106799–106810. https://doi.org/10.1109/ACCESS.2020.3000739https://ieeexplore.ieee.org/document/9110884

Kay, J. (2012). AI and education: Grand challenges. *Intelligent Systems, IEEE, 27*(5), 66–-69. https://doi.org/10.1109/MIS.2012.92 https://www.researchgate.net/publication/260330353_AI_and_Education_Grand_Challenges

Klobas, J. E., Mackintosh, B., & Murphy, J. (2014). The anatomy of MOOCs. In P. Kim (Ed.), *Massive open online courses: The MOOC revolution* (pp. 1–22). Routledge. https://www.researchgate.net/publication/272427367_The_anatomy_of_MOOCs

Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *The International Review of Research in Open and Distributed Learning, 12*(3), 19–38. https://doi.org/10.19173/irrodl.v12i3.882

Kop, R., Fournier, H., & Mak, J. S. F. (2011). A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses. *The International Review of Research in Open and Distributed Learning, 12*(7), 74–93. https://doi.org/10.19173/irrodl.v12i7.1041

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education: A literature review. In A. Peña-Ayala (Ed.), *Learning analytics: Fundamentals, applications, and trends* (pp. 1–-23). Springer International Publishing AG. https://link.springer.com/chapter/10.1007/978-3-319-52977-6_1

Liu, M., Kang, J., Cao, M., Lim, M., Ko, Y., Myers, R., & Schmitz Weiss, A. (2014). Understanding MOOCs as an emerging online learning tool: Perspectives from the students. *American Journal of Distance Education, 28*(3), 147–159. https://doi.org/10.1080/08923647.2014.926145

Lundqvist, K., Liyanagunawardena, T., & Starkey, L. (2020). Evaluation of student feedback within a MOOC using sentiment analysis and target groups. *The International Review of Research in Open and Distributed Learning, 21*(3), 140–156. https://doi.org/10.19173/irrodl.v21i3.4783

Marciniak, R. (2018). Quality assurance for online higher education programmes: Design and validation of an integrative assessment model Applicable to Spanish universities. *The International Review of Research in Open and Distributed Learning, 19*(2), 127–154. https://doi.org/10.19173/irrodl.v19i2.3443

Marin, V., Pereira, T., Sridharan, S., & Rivero, C. (2017). Automated personalized feedback in *Introduction Java Programming* MOOCs. *2017 IEEE 33rd International Conference on Data Engineering* (ICDE; pp. 1259–1270). https://doi.org/10.1109/ICDE.2017.169 https://ieeexplore.ieee.org/document/7930065

McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). *In the open: The MOOC model for digital practice* (Knowledge Synthesis Grants on the Digital Economy) [Grant Report]. Social Sciences & Humanities Research Council of Canada. https://www.oerknowledgecloud.org/archive/MOOC_Final.pdf

McNamara, D., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining massive open online class. *Proceedings of the 8th International Educational Data Mining Society* (pp., 388–391). https://www.researchgate.net/publication/319764516_Language_to_Completion_Success_in_an_Educational_Data_Mining_Massive_Open_Online_Class

Moreno-Marcos, P., Alario-Hoyos, C., Munoz-Merino, P., Estévez-Ayres, I., & Delgado-Kloos, C. (2018). Sentiment analysis in MOOCs: A case study. *IEEE Global Engineering Education Conference (EDUCON;* pp. 1489–1496). https://ieeexplore.ieee.org/document/8363409

Ochando, H. M. P. (2017). Transcultural validation within the English scope of the questionnaire evaluation of variables moderating style of teaching in higher education. CEMEDEPU. *Procedia-Social and Behavioral Sciences, 237*, 1208–1215. https://doi.org/10.1016/j.sbspro.2017.02.191

Onah, D., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: Behavioural patterns. *Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN14)*. https://doi.org/0.13140/RG.2.1.2402.0009 https://www.researchgate.net/publication/273777281_Dropout_Rates_of_Massive_Open_Online_Courses_Behavioural_Patterns

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64. https://www.j-ets.net/collection/published-issues/17_4 https://www.researchgate.net/publication/267510046_Learning_Analytics_and_Educational_Data_Mining_in_Practice_A_Systematic_Literature_Review_of_Empirical_Evidence

Pérez, R. C., Jurado, F., & Villen, A. (2019). Moods in MOOCs: Analyzing emotions in the content of online courses with edX-CAS. *IEEE Global Engineering Education Conference* (EDUCON; pp. 1467–1474). https://doi.org//10.1109/EDUCON.2019.8725107ieeexplore.ieee.org/document/8725107

Shatnawi S., Gaber M. M., & Cocea M. (2014) Automatic content related feedback for MOOCs based on course domain ontology. In E. Corchado, J. A. Lozano, H. Quintián, & H. Yin (Eds.) *Intelligent data engineering and automated learning, IDEAL 2014* (Lecture Notes in Computer Science, Vol. 8669). Springer. https://doi.org/10.1007/978-3-319-10840-7_4

Shatnawi, S., Gaber, M., & Haig, E. (2014). Automatic content related feedback for MOOCs based on course domain ontology (Vol. 8669). *International Conference on Intelligent Data Engineering and Automated Learning*. https://www.researchgate.net/publication/265132990_Automatic_Content_Related_Feedback_for_MOOCs_Based_on_Course_Domain_Ontology

Sindhu, I., Daudpota, S., Badar, K., Bakhtyar, M., Baber, J., & Nurunnabi, M. (2019). Aspect based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access, 7*, 108729–108741. https://doi.org/10.1109/ACCESS.2019.2928872

Su, Y. S., & Lai, C. F. (2021). Applying educational data mining to explore viewing behaviors and performance with flipped classrooms on the social media platform Facebook. *Frontiers in Psychology*, *12*, 653018. https://doi.org/10.3389/fpsyg.2021.653018

Su, Y. S., & Wu, S. Y. (2021). Applying data mining techniques to explore users behaviors and viewing video patterns in converged it environments. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-02712-6

Su, Y. S., Ding, T. J., & Chen, M. Y. (2021). Deep learning methods in internet of medical things for valvular heart disease screening system. *IEEE Internet of Things Journal*. https://ieeexplore.ieee.org/document/9330761

Swinnerton, B., Morris, N., Hotchkiss, S., & Pickering, J. (2016). The integration of an anatomy massive open online course (MOOC) into a medical anatomy curriculum: Anatomy MOOC. *Anatomical Sciences Education, 10*(1), 53–-67. https://doi.org/10.1002/ase.1625https://pubmed.ncbi.nlm.nih.gov/27315160/

Tekin, C., Braun, J., & van der Schaar, M. v. d. (2015). *eTutor: Online learning for personalized education* [Paper presentation]. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brisbane, Australia. https://arxiv.org/abs/1410.3617

Veletsianos, G. (2013). *Learner experiences with moocs and open online learning*. Hybrid Pedagogy Publica. http://learnerexperiences.hybridpedagogy.com

Wen, M., Yang, D., & Rose, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014;* pp. 130–137). https://www.dropbox.com/s/crr6y6fx31f36e0/EDM%202014%20Full%20Proceedings.pdf

Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research, 57*(3), 547-570. https://doi.org/10.1177/0735633118757015

Yang, S. J. H., Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Ogata, H., & Lin, A. J. Q. (2018). Predicting students' academic performance using multiple linear regression and principal component analysis. *Journal of Information Processing, 26*(2), 170-176. https://doi.org/10.2197/ipsjjip.26.170

Yang, T. Y., Brinton, C. G., Joe-Wong, C., & Chiang, M. (2017). Behavior-based grade prediction for MOOCs via time series neural networks. *IEEE Journal of Selected Topics in Signal Processing, 11*(5), 716-728. https://doi.org/10.1109/JSTSP.2017.2700227

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 23, Number 1

February – 2022

An Intelligent Nudging System to Guide Online Learners

M. Elena Rodríguez[1,2], Ana Elena Guerrero-Roldán[1,2], David Baneres[1,2], and Abdulkadir Karadeniz[2,3]
[1]Faculty of Computer Science, Multimedia and Telecommunications, Universitat Oberta de Catalunya; [2]eLearn Center, Universitat Oberta de Catalunya; [3]Open Education Faculty, Anadolu University

Abstract

This work discusses a nudging intervention mechanism combined with an artificial intelligence (AI) system for early detection of learners' risk of failing or dropping out. Different types of personalized nudges were designed according to educational principles and the learners' risk classification. The impact on learners' performance, dropout reduction, and satisfaction was evaluated through a study with 252 learners in a first-year course at a fully online university. Different learners' groups were designed, with each receiving a different set of nudges. Results showed that nudges positively impacted the learners' performance and satisfaction, and reduced dropout rates. The impact significantly increased when different types of nudges were provided. Our research reinforced the role of AI as useful in online, distance, and open learning for providing timely learner support, improved learning experiences, and enhanced learner-teacher communication.

Keywords: artificial intelligence, early warning system, nudges, at-risk learners, online learning

Introduction

Software systems to assist learners and support teachers' tasks in higher education (HE) have evolved in recent years. HE institutions, particularly fully open and distance universities, have shared their vast expertise about using these systems in a range of educational environments (Castro, 2019). When combined with artificial intelligence (AI) techniques, these software systems have become intelligent systems (Chen et al., 2020) capable of analyzing large educational datasets coming from learning management and other university systems (Siemens & Baker, 2012). Inferred knowledge has enabled educators to make decisions based on evidence, thereby impacting education in different dimensions (Chassignol et al., 2018).

AI-based systems improve learner success and retention by enabling early detection and support of online learners at risk of failing or dropping out; these are key concerns in online learning (Grau-Valldosera et al., 2019). To this end, we developed an adaptive intelligent system (called LIS system) with predictive analytics, a progression dashboard, automated nudges, and recommendations based on AI classification algorithms. There has been considerable research in early detection of at-risk learners. Although predictive models and systems have been proposed (Arnold & Pistilli, 2012; Márquez-Vera et al., 2016; Ortigosa et al., 2019; Vasquez et al., 2015), subsequent support for learners is still an open issue. Our work aimed to develop a nudging intervention mechanism in conjunction with an AI-based system to detect at-risk learners early, and to evaluate the system's overall impact on learner performance, dropout rates, and student satisfaction.

Literature Review

Nudges are "interventions that preserve freedom of choice that nonetheless influence people's decisions" (Sunstein, 2015, p. 2). Their effectiveness has been evaluated in health (Bucher et al., 2016), human-computer interaction (Caraban et al., 2019), and across disciplines (Benartzi et al., 2017; Hummel & Maedche, 2019).

In education, recent work (Weijers et al., 2020) has stated that the application of nudging has been sparse, constituting a new research field. Nevertheless, the literature suggested that nudges impact engagement, task completion, and the study of learning resources (Kraft & Rogers, 2015; Martinez, 2014; van Oldenbeek et al., 2019; York et al., 2019). As Mitrovic et al. (2019) noted, nudges foster constructive learning, while Piotrkowicz et al. (2020) discussed the effectiveness of nudges in lifelong e-learning. The systematic review by Damgaard and Nielsen (2018), which included online and distance learning experiences, provided valuable insights: (a) learners appreciate nudges; (b) nudges produce short-term effects; and (c) nudges rarely produce positive effects for all learners.

Overall, these findings suggested that it is better to focus on improving short-end goals that are not behaviors themselves, and that personalized nudges are required. AI allows for such personalization. AI-based systems oriented to support at-risk learners early (Márquez-Vera et al., 2016; Ortigosa et al., 2019; Vasquez et al., 2015) have produced forecasting information, and learners can be nudged through feedforward mechanisms to prevent failure outcomes (i.e., short-term goals). Although there has been discussion about what is considered feedforward (Reimann et al., 2019; Sadler, 2010), it typically refers to future-oriented feedback applied to upcoming assignments. Furthermore, nudges have promoted a learner-teacher relationship that may positively impact learners' satisfaction and learning outcomes

(Ajjawi & Boud, 2018; Eom et al., 2006; Sparks et al., 2015). Therefore, our choice to build a nudging intervention mechanism within the LIS system to complement educational feedback, a cornerstone to support online learners (Martínez-Argüelles et al., 2015), was appropriate.

System Overview

Study Context

All educational activity at the Universitat Oberta de Catalunya (UOC) occurs within its virtual campus. Courses are organized in virtual classrooms attended by teachers. The educational model is learner-centered; it provides all learning resources as well as continuous assessment combined with summative evaluation tailored to each course. There are two types of feedback to support learners—general and personalized. General feedback is addressed to all learners who share a virtual classroom, and is provided by teachers through their blackboard, a communication space where the teacher can post relevant information about the course. After each activity is assessed, each learner also receives personalized feedback, together with her mark.

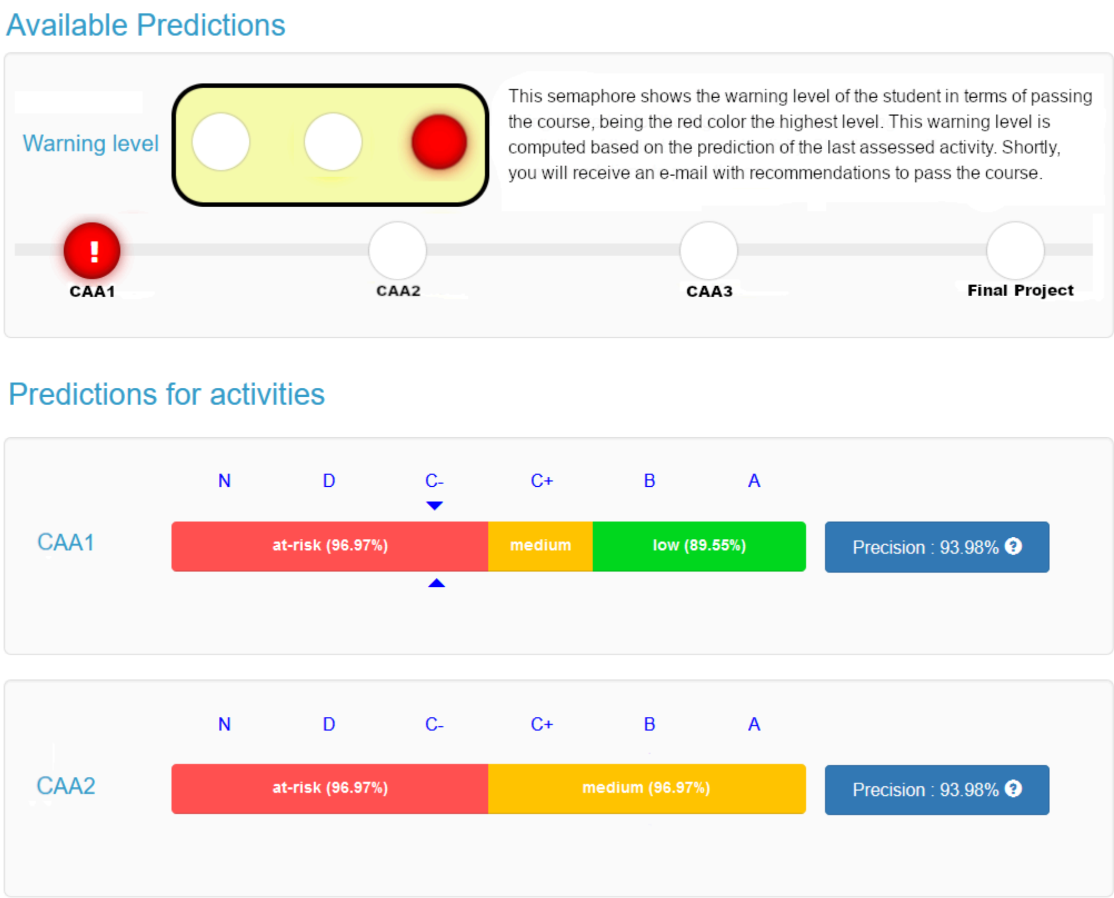
The Early Warning System

As part of the LIS system, an early warning system (EWS) detects learners at risk of failing or dropping out. The EWS uses AI techniques to detect these learners through their grades for the continuous assessment activities (CAA) and each learner's profile. The system considers the number of courses the learner has enrolled in, whether she is a new learner, how many times she has enrolled in the course, and her grade point average.

The predictive model is trained with anonymized data from past learners. The predictive model for a course consists of as many submodels as CAAs in the course where the most suitable classification algorithm is applied in terms of accuracy, from among decision tree (DT), k-nearest neighbors (KNN), support vector machine (SVM), and naive Bayes (NB). For each CAA, a prediction is issued. Using the submodel associated with the CAA being analyzed, a simulation detects the minimum grade for a learner to obtain in the next CAA in order to avoid risk of failing. The submodel uses the learners' profile and her earlier CAA grades to simulate all possible grades for the next CAA, thereby identifying the grade that will change the prediction from fail to pass. This minimum grade is compared with the grade the learner finally obtains for the CAA. Such comparison generates a risk warning level (high, moderate, low) using a green-amber-red semaphore, similar to Arnold and Pistilli (2012). Prediction is personalized because it depends on her profile and previous CAA performance. Each CAA is qualitatively graded (i.e., A, B, C+, C-, or D); grades from C+ to A indicate a pass. The grade N means the learner did not deliver the CAA. Figure 1 shows the progression dashboard for a learner who has received the warning level classification for the first CAA.

Figure 1

Learner's Progression Dashboard



When the learner is notified about a risk warning level (e.g., high risk of failure because she obtained a C- grade), the risk level distribution for the subsequent CAA is also adjusted. Thus, the learner knows in advance which grades she must obtain in the next CAA in order to pass the course (according to Figure 1, she needs a minimum C+ grade). Both teachers and learners are notified of an at-risk classification, though the main focus is on learners likely to fail, in order to apply interventions to revert the at-risk situation. An in-depth analysis of the EWS is available at Baneres et al. (2020).

The Intervention Mechanism

Table 1 summarizes the types of messages and nudges supported by our intervention mechanism, as outlined in the classification system proposed by Damgaard and Nielsen (2018). Table 1 also includes information on personalization levels and suggested content. The nudges supported by the intervention mechanism are a consequence of an in-depth analysis of the continuous assessment strategies at UOC and semi-structured interviews with expert teachers.

Messages were linked to each CAA in the course. Two events deal with developing the CAA, while two events are associated with assessing it. The messages were triggered automatically by the system on the teacher's behalf when certain events and conditions hold, and messages were sent by e-mail from the teacher to the learners.

Informational messages (I1) were associated with beginning the CAA. All learners received the same message (i.e., low personalization is required) containing information about the CAA's objectives, learning outcomes, and the available learning resources. I1 also highlighted the importance of good planning and urged learners to develop this skill.

Messages could be scheduled, so a few days after the CAA started, learners could be sent a reminder (R1) that the CAA had begun. The degree of personalization was medium because the teacher could set receipt of R1 to only those who have not accessed the virtual classroom since the CAA began. Learners having trouble were prompted to contact the teacher for individualized assistance. Similarly, R2 was sent when the CAA deadline approached. Only learners who had not submitted the CAA received R2, and the teacher specified that the R2 trigger occurred a set number of days before the R2 deadline.

Table 1

Nudges Supported by Intervention System

CAA stage	Event	Message type	Nudge type	Personalization level	Suggested content
CAA development	Beginning	Informational (I1)	Goal setting Informational	Low	CAA in context Available learning resources Advice for goal setting and planning Importance of communication spaces
		Reminder 1 (R1)	Reminder Social belonging Assistance	Medium	Warning that CAA has started Encourage learner to participate or be aware of communication spaces Boost the learner to contact the teacher in case of problems
	Submission	Reminder 2 (R2)			Warning that CAA deadline is close Submission requirements

CAA assessment	Solution	Feedback message 1 (FM1)	Informational Goal setting Assistance Extrinsic motivation	Medium	Advice about the CAA solution and alternative solutions Suggestion to improve planning Provide key concepts and competencies for the upcoming CAA Provide learning resources Options to pass the course Teacher assistance
	Mark	Feedback message 2 (FM2)	Informational Assistance	High	Predictive statement, warning level, and options to pass (feedforward) Teacher assistance

FM1 was linked to the teacher's published CAA solution, and it supported medium personalization—the message changed depending on whether the learner submitted the CAA. It was possible to distinguish between learners who simply failed to submit the last CAA and those who had not submitted two or more consecutive CAA. Learners who submitted the CAA received information about using the solution to enhance their learning as well as alternative solutions. Learners were encouraged to compare their answers to the teacher's solution, ask questions, and review their planning.

Learners who had not submitted the CAA were advised about the key concepts and competencies to succeed in the upcoming CAA as well as the learning resources they should study; they were urged to ask for individualized assistance. They may also have received extrinsic motivation. Learners who had not submitted more than one CAA received information about alternatives to achieving a passing grade (e.g., mandatory CAA, an examination when the semester ends), and they could also unsubscribe from the messaging system if they wished.

FM2 was sent when the CAA was graded; it explained the prediction issued and pushed learners to consult their dashboard to improve their warning level in the next CAA. FM2 varied depending on the learner's warning level, allowing for high personalization. Learners at low risk received a congratulation message. Learners could be classified as at medium risk for three reasons, and the message differed in each case. First, the learner passed the CAA but with a grade lower than the minimum grade suggested by the EWS. Second, the EWS model inaccurately predicted that she was not at risk, in which case the EWS indicated the model's lack of accuracy and warned her about potential future problems. Third, the

EWS inaccurately predicted that the learner was at risk. Learners at high risk received different messages, depending on whether or not they submitted the CAA (i.e., the system distinguished between one or several consecutive non-submitted CAAs). Learners who submitted but failed and were at risk, and for whom the prediction is accurate, received a message that positively valued their effort and offered personalized assistance.

The intervention mechanism was able to adapt. Teachers were able to choose which messages to send depending on course characteristics. Some messages could be combined. For example, I1 could be integrated into FM1 when CAAs were related. Similarly, FM1 could be integrated into FM2 when the CAA solution and grades were published at close to the same time. However, it was mandatory to provide FM2, as it dealt with the issued predictions, which were probability estimations. To avoid discouragement or overconfidence, which could have negatively impacted their performance, learners must have understood the uncertainty level.

Methodology

Research Questions

We proposed that higher personalization at the appropriate time in a course positively impacted learners' performance and satisfaction. Using a nudging intervention mechanism combined with an EWS supported both these requirements. Therefore, we identified three research questions:

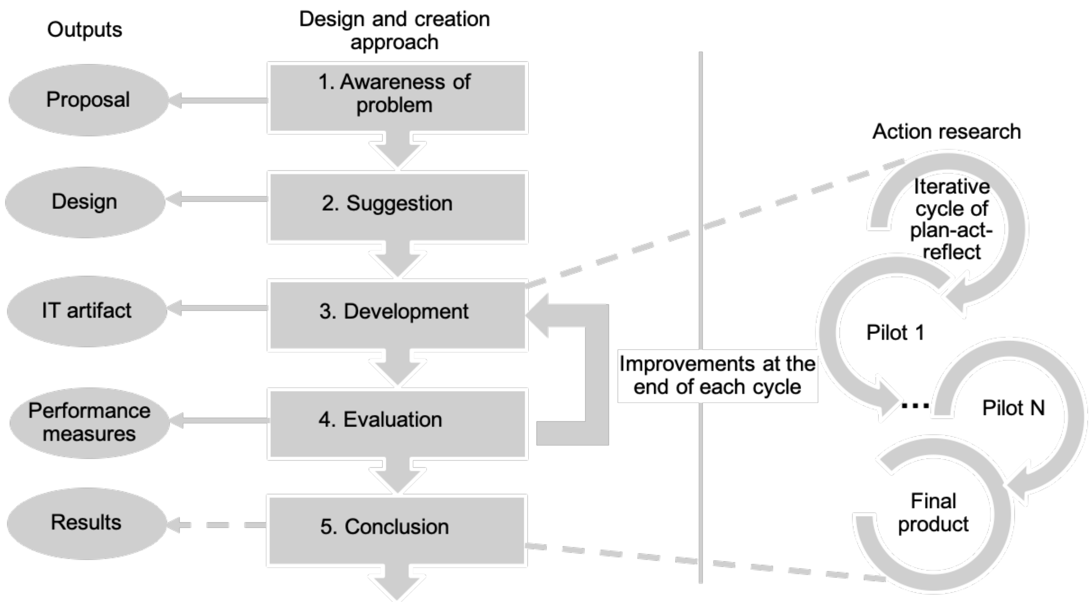
1. Is there an impact on learners' performance when different nudge types are received?
2. Is there an impact on the dropout rate during the continuous assessment when different nudge types are received?
3. What are learners' opinions about the usefulness, engagement, and their own mood based on the different nudge types received?

Research Method

LIS system development followed a mixed research methodology (see Figure 2) that combined an action research methodology with a design and creation approach (Oates, 2006).

Figure 2

Research Method



Once a problem was detected and shared (i.e., learners' at-risk identification and support was required), an artifact solution (i.e., the LIS system) was suggested. Next, the artifact was gradually implemented and tested in real scenarios following an iterative cycle of plan-act-reflect. After each cycle, an evaluation was done according to performance measures. Depending on the results, changes in the artifact were introduced, causing a new cycle until the final artifact was obtained. The research we present constituted a cycle (see Figure 3) conducted in the second semester of the 2019–2020 academic year, where the nudging intervention mechanism was tested.

Participants

The study participants were learners from the computer science bachelor's degree. Participants were enrolled in the first-year *Computer Fundamentals* course where they learned to analyze and synthesize digital circuits and developed an understanding of the underlying computer architecture. Learning resources for the course were text-based and multimedia materials. The continuous assessment model comprised three assessment activities (CAA1, CAA2, and CAA3) as well as the final project (FP). A face-to-face exam at the end of the semester complemented the continuous assessment. Although the activities were assessed using a qualitative scale, grades (Gr) were transformed at the end of the semester to numerical values (A: 9/10, B: 7/8, C+: 5/6, C-: 3/4, and D: 0/1/2). The final mark (FM) was computed as follows:

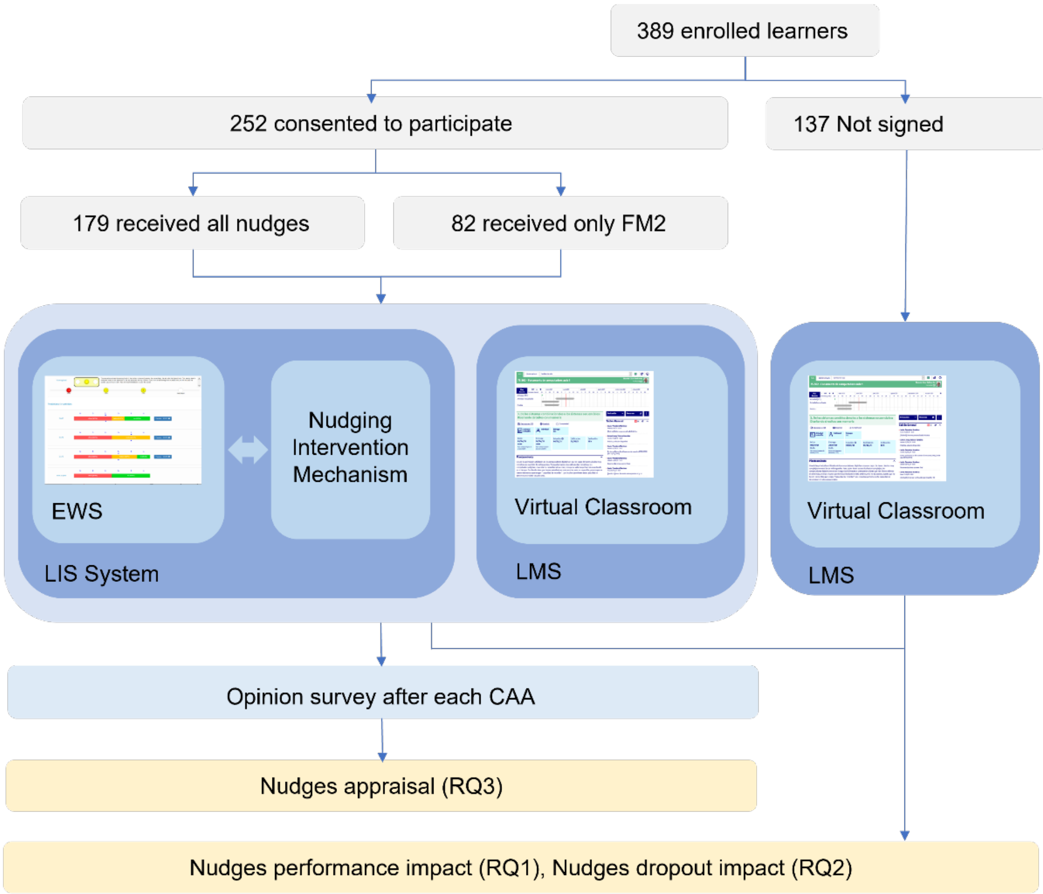
$$FM = MAX\,(10\%Gr_{CAA1} + 10\%Gr_{CAA2} + 10\%Gr_{CAA3} + 35\%Gr_{FP} + 35\%Gr_{EXAM},\ 50\%Gr_{FP} + 50\%Gr_{EXAM})$$

As we can observe in the previous formula, it was possible for learners to pass the course without performing CAA1, CAA2, and CAA3, but the final project (FP) and the exam were mandatory. Teachers conditioned learners must have reached a minimum grade of 4 on both the FP and the exam to pass; the grading system went from 0 to 10, with 5 as the lowest passing grade.

Computer Fundamentals was a suitable course for our analysis because there was a low academic success ratio (40% to 50% from enrollment), mainly due to students who dropped out. Although it was possible to pass the course without performing some CAA, teachers knew that learners who did not perform them had difficulties. Previous research (Rodríguez et al., 2019) concluded that dropout rates in the course were related to failing or not submitting the CAA. Several factors have affected that. First, learners were required to manage their academic work plus professional and family commitments. As well, learners reported that their course workload sometimes meant they were faced with similar deadlines for multiple CAA. Finally, learners encountered difficulties in the course content, CAA perceived difficulty and length, and the appropriateness of learning resources. These factors were even more relevant because it was a first-year course, and many learners were new to online education.

Figure 3

Research Procedure



The risk level classification depended on the accuracy of the predictive model available through the EWS, which included as many submodels as CAAs in the course. Table 2 shows the accuracy of the submodels for *Computer Fundamentals*. The metrics were: (a) the number of at-risk learners correctly identified (TP); (b) the number of non-at-risk learners correctly identified (TN); (c) the number of at-risk learners not correctly identified (FP); (d) the number of non-at-risk learners not correctly identified (FN); (e) the global accuracy of the model (ACC); (f) the accuracy when detecting at-risk learners (TPR true positive rate); (g) the accuracy when distinguishing non-at-risk learners (TNR, true negative rate); (h) the F-score ($F_{1.5}$) made up of a harmonic mean of the true positive value (precision) and the TPR (recall) that weighted correct at-risk identification; (i) and the selected classification algorithm (classifier). The accuracy of detecting non-at-risk learners (TNR) started at 77.72% and reached a value of 94.69% in the last activity. Detecting at-risk learners (TPR) started at 71.78%, but it reached a similar value to the TNR in the last activity, namely 93.77%. In most cases, learners received the right nudges regarding their actual failing risk with this level of accuracy.

Table 2

Performance of the Predictive Model to Identify Potential Course Failure

Submodel	TP	FP	TN	FN	ACC (%)	TNR (%)	TPR (%)	$F_{1.5}$ (%)	Classifier
Pr_{CAA1}	96	46	117	28	74.22	77.72	71.78	74.30	DT
Pr_{CAA2}	192	74	279	15	84.11	92.75	79.04	83.32	DT
Pr_{CAA3}	184	29	324	23	90.71	88.89	91.78	92.27	SVM
Pr_{FP}	196	22	331	11	94.11	94.69	93.77	94.68	KNN

We initially designed two learners' groups to analyze the research questions, each receiving different nudge sets (see Figure 3). The UOC committee for research ethics required that learners consent to participate in any study following the European General Data Protection Regulation (https://gdpr-info.eu/). Once consent was received, the LIS system processed the learners' anonymized data. Due to this, a third learner group was included: the learners who declined to participate. Each of the three learner groups received the following nudges (see Table 1):

1. All nudges: Learners who signed the consent and received I1, R1, R2, FM1, and FM2.

2. FM2 only: Learners who signed the consent and received only FM2.

3. Not signed: Learners who did not sign the consent and did not receive a nudge.

Of the 389 enrolled learners in *Computer Fundamentals*, 170 (43.70%) learners signed the consent and were placed in the first group, 82 (21.07%) signed the consent and were placed in the second group, and 137 (35.21%) learners who did not sign the consent were assigned to the third group.

Instruments

Three instruments were used for collecting data. First, quantitative data about learners' performance and dropout rates were obtained from the institutional information systems. Second, data concerning learners' risk came from the EWS. All data were stored in comma-separated values format. R language was used to merge and analyze the datasets. For the first research question, statistical significance analysis of performance was done using the unpaired two-sample Wilcoxon test due to the non-normal distribution of the final mark (Kruskal, 1957). Descriptive analysis showed the difference in median, mean, standard deviation, minimum, and maximum values. For the second research question, the dropout rate difference for each learners' group was analyzed. Finally, for qualitative data, a questionnaire embedded into the EWS was used. Thus, the third research question was supported by analyzing Likert scale average values in this opinion survey.

Results

Research Question One: The Impact of Different Nudge Types on Learners' Performance

First, we analyzed each group's performance, and then, the groups' statistical significance. Performance data were based on the learners' final marks. The groups were filtered by removing learners who did not submit any CAA. Such learners dropped out of the course before submitting CAA1. Many of them did not start the course, so including them would have skewed the findings and conclusions.

Table 3 summarizes the participants' demographic information. There were 157 participants after the filtering process. Removing learners who dropped out from the beginning (those who did not start the CAA) mainly impacted the not signed group (i.e., 27.01%). Their impact on other groups was significantly less (i.e., 7.65% on all nudges group and 7.32% on FM2 only group). Concerning gender distribution, there was a gender imbalance, consistent with women's minority presence in science, technology, engineering, and mathematics (Barr, 2014). Finally, participants' ages ranged from 21 to 45 years in all groups, and it did not influence the participation in the study.

Table 3

Demographic Information

Participant category	All nudges		FM2 only		Not signed	
	n	Participant rate	*n*	Participant rate	*n*	Participant rate
Group						
Filtered	157	92.35%	76	92.68%	100	72.99%
CAA not started	13	7.65%	6	7.32%	37	27.01%
Total by group	170		82		137	
Gender						
Male	133	84.71%	68	89.47%	94	94.00%
Female	24	15.29%	8	10.53%	6	6.00%
Total by group	157		76		100	
Age range in years						
(18–20)	6	3.82%	1	1.32%	1	1.00%
(21–25)	38	24.20%	23	30.26%	35	35.00%
(26–30)	33	21.02%	14	18.42%	22	22.00%
(31–35)	26	16.56%	12	15.79%	11	11.00%
(36–40)	24	15.29%	12	15.79%	12	12.00%
(41–45)	18	11.46%	7	9.21%	7	7.00%
(46–50)	9	5.73%	6	7.89%	7	7.00%
(51–55)	3	1.91%	1	1.32%	5	5.00%

>55	0	0.00%	0	0.00%	0	0.00%
Total by group	157		76		100	

Table 4 summarizes the descriptive statistics for the groups by age distribution. The number of participants after the filtering process (filtered), as well as the median, mean, standard deviation (*SD*), minimum (min.), and maximum (max.) values of the final marks are shown. Each group was represented in all mark categories, but the median and mean were significantly higher when learners received nudges. Receiving all nudge types had even more impact on the final mark median. The grade distribution did not follow a normal distribution. There was a high dispersion from the minimum value (zero) to the maximum value (10), as indicated by the *SD* variability. However, comparing the median and the mean indicated that a large number of learners passed the course (grade equal to or higher than five) with higher grades when they received all nudge types. When receiving FM2 only, metrics values were similar. Finally, learners who did not receive nudges tended to fail the course irremediably. Regarding age distribution, we observe the same tendency. Learners in all age ranges who received more nudges improved their performance to a considerable extent.

Table 4

Final Mark Distribution: Descriptive Statistics for Each Group

Age range in years	Filtered	Median	Mean	*SD*	Min.	Max.
All nudges						
(18–20)	6	8.50	6.48	3.57	0.00	9.90
(21–25)	38	7.50	5.74	3.83	0.00	10.00
(26–30)	33	7.20	6.47	3.64	0.00	10.00
(31–35)	26	8.00	6.59	3.60	0.00	10.00
(36–40)	24	7.80	5.75	4.31	0.00	10.00
(41–45)	18	7.25	6.01	3.85	0.00	10.00
(46–50)	9	7.30	6.84	2.87	0.00	9.90
(51–55)	3	7.00	5.67	5.13	0.00	10.00
>55	0	--	--	--	--	--
Total by group	157	7.60	6.16	3.73	0.00	10.00
FM2 only						
(18–20)	1	2.00	2.00	--	2.00	2.00
(21–25)	23	5.90	5.16	3.70	0.00	10.00
(26–30)	14	5.10	4.76	3.49	0.00	10.00
(31–35)	12	7.20	5.74	4.17	0.00	10.00
(36–40)	12	6.10	5.45	3.84	0.00	10.00
(41–45)	7	5.60	4.50	4.40	0.00	10.00

(46–50)	6	8.45	7.35	3.69	0.00	9.90
(51–55)	1	7.30	7.30	--	7.30	7.30
>55	0	--	--	--	--	--
Total by group	76	5.80	5.26	3.72	0.00	10.00
			Not signed			
(18–20)	1	0.00	0.00	--	0.00	0.00
(21–25)	35	0.00	2.82	3.30	0.00	9.10
(26–30)	22	0.00	2.15	3.44	0.00	9.30
(31–35)	11	7.30	5.36	4.35	0.00	9.80
(36–40)	12	7.65	4.95	4.42	0.00	9.90
(41–45)	7	2.00	4.03	4.47	0.00	9.30
(46–50)	7	2.45	3.95	4.27	0.00	9.60
(51–55)	5	0.00	1.87	3.75	0.00	7.5
>55	0	--	--	--	--	--
Total by group	100	0	3.32	3.83	0.00	9.90

The unpaired two-sample Wilcoxon test was used to check the statistical significance of the distribution of the improvement in final marks. The null hypothesis was that the scores were worse or equal when more nudges were received. Table 5 shows the comparison among all groups.

Table 5

Results of the Unpaired Two-Sample Wilcoxon Test on Final Mark Distribution

Group	Group comparison	*p*-value	Significance [a]	Hypothesis
All nudges	FM2 only	0.021	*	Reject
All nudges	Not signed	$2.3e^{-08}$	****	Reject
FM2 only	Not signed	0.00028	***	Reject

Note. [a] Significance: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Although the hypothesis was rejected in all cases, there were different significance levels. The comparison with the learners who did not sign was clearly significant (i.e., *p*-value $2.3e^{-08}$ compared to the all nudges group and 0.00028 compared to the FM2 only group). Thus, obtaining some additional nudges during the CAA positively impacted learners' performance. When comparing different nudge types (i.e., all nudges vs. FM2 only), the significance level was lower with a *p*-value of 0.021. However, there was still some significance when receiving all nudge types. The results were consistent with the descriptive statistics of Table 4. Receiving FM2 gave learners information about the assessed CAA and their risk level only. However, receiving additional nudges enlightened learners with information about competencies to acquire in the next CAA, skills needed from the previous CAA, and reminders about the next submission. Figure 4 summarizes the results with a notched box-and-whisker plot.

Research Question Two: Impact of Receiving Different Nudge Types on Dropout Rates

We analyzed how the dropout rate was affected depending on the nudge types provided. We excluded learners who did not submit any CAAs and the FP. Consequently, Table 6 does not summarize dropout for CAA1 because such learners were excluded from the analysis. As a result, Table 6 reports the dropout rate for each CAA and the FP showing the number of participants (filtered), dropped out learners (*n*), and percentages of learners within the group (%).

Figure 4

Box-And-Whisker Plot of the Final Mark Distribution With Corresponding p-Value of the Unpaired Two-Sample Wilcoxon Test

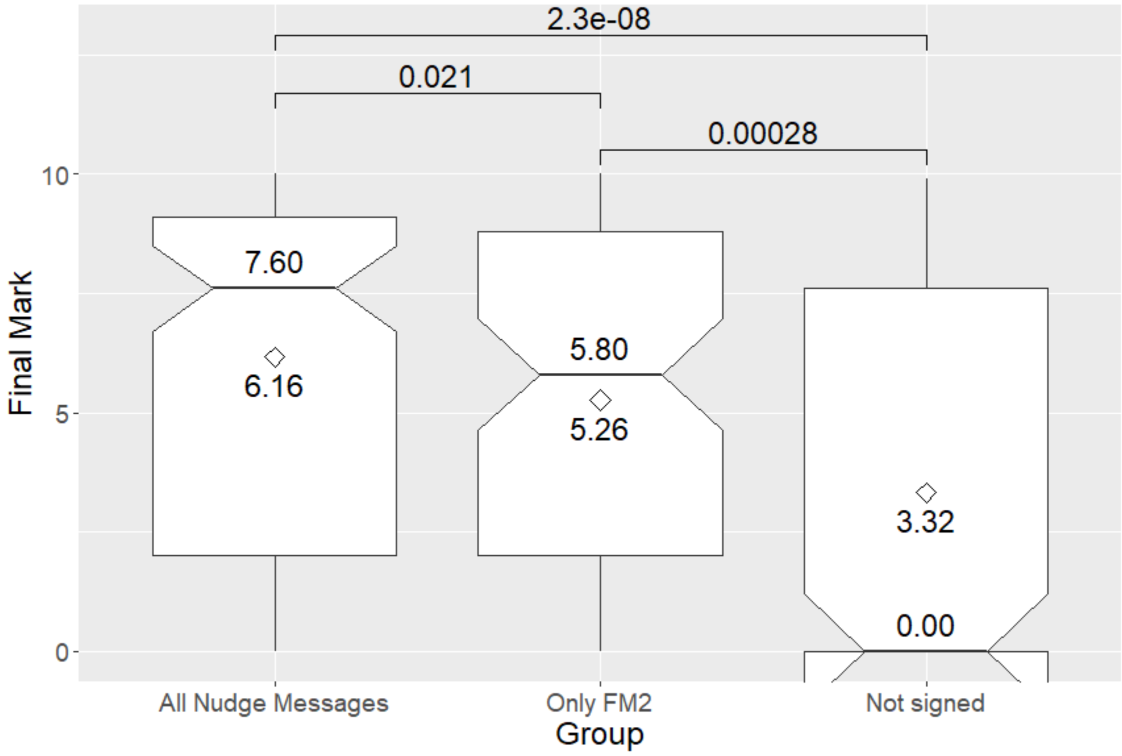


Table 6

Dropout Rate for Each Group and Assessment Activity

Group	Filtered	CAA2		CAA3		FP	
		n	%	*n*	%	*n*	%
All nudges	157	14	8.92	26	16.56	45	28.66
FM2 only	76	6	7.89	13	17.11	25	32.89
Not signed	100	26	26.00	38	38.00	57	57.00

The dropout rate was higher for learners who did not sign the consent. By the end of the semester, more than half of them had dropped out. *Computer Fundamentals* is a first-year course; many learners were new to university studies and to online learning. These factors greatly influenced the dropout rate, which was 35.21% on average. However, at the end of the semester, the dropout rate for learners who received nudges (FM2 only or the complete set) was lower than the average.

Research Question Three: Learners' Opinions About Usefulness, Engagement, and Their Mood Regarding Different Types of Nudges

Once each CAA was graded, the risk level assigned, and the prediction for the upcoming CAA made available, learners were prompted to answer a short opinion questionnaire. Answers were based on a Likert scale from 1 (*strongly negative*) to 5 (*strongly positive*). There were three questions: (a) Do you think the received messages are useful? (b) Are you going to continue the course? and (c) What is your mood after receiving the messages? Since the survey was embedded in the EWS, we were able to associate the learners' responses with their group. Table 7 summarizes the results and shows the Likert scale average values on the three questions for each CAA and the FP.

Table 7

Learners' Opinion About the Nudges: Usefulness, Engagement, and Their Mood

Group	Usefulness	Engagement	Mood	Responses
		CAA1		
All nudges	4.00	4.04	3.88	74
FM2 only	3.82	4.00	3.73	38
		CAA2		
All nudges	3.85	3.89	3.83	91
FM2 only	3.74	3.84	3.82	50
		CAA3		
All nudges	4.03	4.05	3.86	87
FM2 only	4.03	4.00	3.90	39
		FP		
All nudges	3.90	4.04	3.89	85
FM2 only	3.68	3.88	3.71	41

Note. Sample sizes: all nudges (n = 157), FM2 only (n = 76).

All learners considered that the nudges were helpful for their learning process, with an appraisal higher than 65% on average and reaching values near 75%. Learners who received all nudge types provided a higher appraisal. A similar effect was observed in engagement with a value higher than 70%. Learners who received more nudges expected to continue the course with a slightly higher value. Finally, learners considered they had a positive mood during the course with a value higher than 65% on average and higher appraisals when more nudges were received.

Discussion

Concerning the first research question, learners performance improved when more nudge types were received. The statistical significance between learners who did not sign the consent and received no nudges, and those who did, was high (i.e., $p \leq 0.001$). Learners who did not sign the consent received only a final grade, the CAA solutions in the virtual classroom, and general feedback through the teacher's blackboard, without any personalization. Each learner needed to reflect on her mistakes from the CAA solution all on her own and perform this reflection on time. It was difficult for her to know her likelihood of passing. A learner who agreed to be in the study also needed to carry out this reflection, but nudges helped her to do this and set the appropriate time in which to carry out the reflection. The groups who received all nudges or FM2 only received messages with a high degree of personalization. FM2 in particular had a large impact on a learner's performance because it helped her know her place in the course and where to go next. As a feedforward message, it provided learner assistance on how to address the CAAs thereafter. It also gave a backward view of her achievement in past CAAs and a forward view of her likelihood of passing. Despite the discussion about what is considered feedforward (Reimann et al., 2019; Sadler, 2010), its value for "focusing attention on the potential for uptake of information and the necessity of action" is clear (Reimann et al., 2019, p. 10).

When comparing groups of learners who signed the consent, performance was still significant in the group who received all nudges. Results were consistent with the literature. Reminders and informational nudges (Martinez, 2014) enhanced performance and completion rates. Furthermore, when learners received more nudges, they improved their performance significantly in all age ranges with a remarkable result: performance was better by learners aged 31 to 40 years. Research in online settings (Cheung & Kan, 2002; Didia & Hasnat, 1998) has also observed that maturity, combined with previous online learning experience, improved self-regulation and impacted performance.

In terms of research question two, there was a significant reduction in dropout rates in the groups that received all nudges and those who received FM2 only. These learners felt better supported and guided as a result of the teacher's recommendations. Learners who did not sign the consent may have felt alone. Only proactive learners used the different communication channels (i.e., the virtual classroom forums or the teacher's e-mail). The big difference was in terms of who started the possible dialogue (Ajjawi & Boud, 2018). For those who did not sign the consent, it was always the learner who initiated dialogue. For students who received nudges, these messages opened the opportunity to reply to the teacher and create a teacher-learner relationship. Our results were consistent with the literature; meaningful teacher-learner relationships created supportive learning environments (Sparks et al., 2015), and promoted self-efficacy and motivation. We cannot underestimate the efficacy of learners receiving messages in their e-mail. Such messages signalled that some action was expected of the learner at a specific time. Learners who did not consent to participate did not receive this signal. They needed to be proactive and access the virtual classroom frequently to be aware of what was going on. Otherwise, they ran the risk of reacting too late, which partially explains why online learners have tended to concentrate their efforts in courses where they have better performance (Grau-Valldosera et al., 2019). Finally, the results showed a significant decrease in the dropout rate among the groups that received all the nudges and the group that received FM2 only. In the literature, some have argued that reminders significantly impact task completion and engagement (Kraft & Rogers, 2015; York et al., 2019), and informational nudges about competencies to address current activity are necessary (Martinez, 2014) in order to encourage learners to revisit previous learning resources and activities.

Regarding the third research question, our results showed that learners were satisfied with the nudges they received. As Eom et al. (2006) claimed, the teacher's personalized messages impacted satisfaction. Martínez-Argüelles et al. (2015) found a relationship between motivation and mood with receipt of personalized information from the teacher. A similar effect was observed in engagement: learners who received more nudges were more likely to expect to continue the course. Finally, learners' mood was appraised more positively when more nudges were received. Higher values were obtained on usefulness and engagement, while slightly lower values were obtained regarding learners' mood but still were above the average. Thus, the learners' opinions about usefulness, engagement, and mood were positive.

Finally, we note some research limitations. Learners decided to participate in the pilot, inducing an auto-selection bias due to the institutional ethical requirements. These were usually the most engaged learners, and their performance is typically better. There is also a gender bias inherent to the course. Auto-selection mainly affected the first and second research questions, while mortality bias (i.e., learners who discontinued using the system and did not answer the opinion survey) affected the third research question.

Conclusions

Our contribution is twofold. First, we present a nudging intervention mechanism combined with an EWS based on AI techniques. Teachers choose which nudge types are appropriate according to educational principles and also when to send them. The nudges are personalized according to learners' risk level and profile, and learners can be nudged with feedforward to prevent a failure outcome. As far as we know, few studies have focused on feedforward nudges; most studies have focused on automatic messages with marks (Clarizia et al., 2018) rather than on nudges to encourage learners, or just on detecting at-risk learners (Vasquez et al., 2015). Our intervention mechanism automatically manages nudges based on the EWS predictions and risk classification.

Second, we study the nudging intervention mechanism in a real online educational setting. The research questions allowed us to analyze their usefulness and effectiveness. Results suggest that nudges positively impact learners' performance and satisfaction. Moreover, their performance and satisfaction increase when more nudge types are sent.

Our findings have a significant impact on online, distance, and open learning practice, reinforcing the role of AI in extracting relevant information from datasets in order to enhance the teaching-learning process. The benefits of our approach are diverse: timely learner support and guidance, better learning experience, personalization, and effective learner-teacher communication. Our experience shows that this approach can coexist with other available feedback mechanisms.

Nevertheless, our intervention mechanism can achieve even better personalization levels. In future studies, the EWS could make better use of learners' data by detecting new learner classes with particular problems (e.g., dropout, self-regulation, special needs), thereby improving the efficacy of nudges. For example, repeater and novice learners have different needs compared to high-performance learners. A deeper qualitative analysis, including interviews and focus groups, should be performed to better understand learners' appraisal of nudges' effectiveness. Finally, a longitudinal study to analyze learner cohorts is required to see if the results persist across semesters.

Acknowledgments

This research was funded by the eLearn Center at Universitat Oberta de Catalunya through the project New Goals 2018NG001 LIS: Learning Intelligent System.

References

Ajjawi, R., & Boud, D. (2018). Examining the nature and effects of feedback dialogue. *Assessment and Evaluation in Higher Education*, *43*(7), 1106–1119. https://doi.org/10.1080/02602938.2018.1434128

Arnold, K. E., & Pistilli, M. D. (2012, April 29). Course signals at Purdue: Using learning analytics to increase student success. In S. Dawson, C. Haythornthwaite, S. B. Shum, D. Gasevic & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330666

Baneres, D., Rodríguez, M. E., Guerrero-Roldán, A. E., & Karadeniz, A. (2020). An early warning system to detect at-risk students in online higher education. *Applied Sciences (Switzerland)*, *10*(13), 4427. https://doi.org/10.3390/app10134427

Barr, V. (2014, December 1). *Women in STEM, women in computer science: We're looking at it incorrectly*. Blog@CACM. https://cacm.acm.org/blogs/blog-cacm/180850-women-in-stem-women-in-computer-science-were-looking-at-it-incorrectly/

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, *28*(8), 1041–1055. https://doi.org/10.1177/0956797617702501

Bucher, T., Collins, C., Rollo, M. E., McCaffrey, T. A., De Vlieger, N., Van der Bend, D., Truby, H., & Perez-Cueto, F. J. A. (2016). Nudging consumers towards healthier choices: A systematic review of positional influences on food choice. *British Journal of Nutrition*, *115*(12), 2252–2263. https://doi.org/10.1017/S0007114516001653

Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P. (2019, May 2). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (article 503). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300733

Castro, R. (2019). Blended learning in higher education: Trends and capabilities. *Education and Information Technologies*, *24*(4), 2523–2546. https://doi.org/10.1007/s10639-019-09886-3

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial intelligence trends in education: A narrative overview. *Procedia Computer Science*, *136*, 16–24. https://doi.org/10.1016/j.procs.2018.08.233

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cheung, L. L. W., & Kan, A. C. N. (2002). Evaluation of factors related to student performance in a distance-learning business communication course. *Journal of Education for Business*, *77*(5), 257–263. https://doi.org/10.1080/08832320209599674

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018, September 23). Chatbot: An education support system for student. In A. Castiglione, F. Pop, M. Ficco & F. Palmieri (Eds.), *Cyberspace Safety and Security, CSS 2018* (Lecture Notes in Computer Science, Vol. 11161, pp. 291–302). Springer. https://doi.org/10.1007/978-3-030-01689-0_23

Damgaard, M. T., & Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review*, *64*, 313–342. https://doi.org/https://doi.org/10.1016/j.econedurev.2018.03.008

Didia, D., & Hasnat, B. (1998). The determinants of performance in the university introductory finance course. *Financial Practice & Education*, *8*(1), 102–107.

Eom, S. B., Wen, H. J., & Ashill, N. (2006). The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. *Decision Sciences Journal of Innovative Education*, *4*(2), 215–235. https://doi.org/10.1111/j.1540-4609.2006.00114.x

Grau-Valldosera, J., Minguillón, J., & Blasco-Moreno, A. (2019). Returning after taking a break in online distance higher education: from intention to effective re-enrollment. *Interactive Learning Environments*, *27*(3), 307–323. https://doi.org/10.1080/10494820.2018.1470986

Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, *80*, 47–58. https://doi.org/10.1016/j.socec.2019.03.005

Kraft, M. A., & Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review*, *47*, 49–63. https://doi.org/10.1016/j.econedurev.2015.04.001

Kruskal, W. H. (1957). Historical notes on the Wilcoxon Unpaired Two-Sample Test. *Journal of the American Statistical Association*, *52*(279), 356–360. https://doi.org/10.1080/01621459.1957.10501395

Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, *33*(1), 107–124. https://doi.org/10.1111/exsy.12135

Martínez-Argüelles, M. J., Plana, D., Hintzmann, C., Batalla-Busquets, J. M., & Badia, M. (2015). Usefulness of feedback in e-learning from the students' perspective. *Intangible Capital*, *11*(4), 627–645. https://doi.org/10.3926/ic.622

Martinez, I. (2014). *The effects of nudges on students' effort and performance: Lessons from a MOOC* (EdPolicyWorks Working Paper Series No. 19). https://education.virginia.edu/sites/default/files/files/EdPolicyWorks_files/19_Martinez_Lessons_from_a_MOOC.pdf

Mitrovic, A., Gordon, M., Piotrkowicz, A., & Dimitrova, V. (2019, June 21). Investigating the effect of adding nudges to increase engagement in active video watching. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren & R. Luckin (Eds.), *Artificial Intelligence in Education, AIED 2019* (Lecture Notes in Computer Science, Vol. 11624, pp. 320–332). Springer.

https://doi.org/10.1007/978-3-030-23204-7_27

Oates, B. J. (2006). *Researching information systems and computing*. SAGE.

Ortigosa, A., Carro, R. M., Bravo-Agapito, J., Lizcano, D., Alcolea, J. J., & Blanco, Ó. (2019). From lab to production: Lessons learnt and real-life challenges of an early student-dropout prevention system. *IEEE Transactions on Learning Technologies*, *12*(2), 264–277. https://doi.org/10.1109/TLT.2019.2911608

Piotrkowicz, A., Dimitrova, V., Hallam, J., & Price, R. (2020, July 14). Towards personalisation for learner motivation in healthcare: A study on using learner characteristics to personalise nudges in an e-learning context. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, (pp. 287–292). https://doi.org/10.1145/3386392.3399290

Reimann, N., Sadler, I., & Sambell, K. (2019). What's in a word? Practices associated with 'feedforward' in higher education. *Assessment and Evaluation in Higher Education*, *44*(8), 1279–1290. https://doi.org/10.1080/02602938.2019.1600655

Rodríguez, M. E., Guerrero-Roldán, A. E., Baneres, D., & Karadeniz, A. (2019, November). Towards an intervention mechanism for supporting learners performance in online learning. *Proceedings of the 12th Annual International Conference of Education, Research and Innovation, ICERI 2019* (pp. 5136–5145). Seville, Spain. https://doi.org/10.21125/iceri.2019.1244

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, *35*(5), 535–550. https://doi.org/10.1080/02602930903541015

Siemens, G., & Baker, R. S. J. D. (2012, April 29). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330661

Sparks, C., Dimmock, J., Whipp, P., Lonsdale, C., & Jackson, B. (2015). "Getting connected": High school physical education teacher behaviors that facilitate students' relatedness support perceptions. *Sport, Exercise, and Performance Psychology*, *4*(3), 219–236. https://doi.org/10.1037/spy0000039

Sunstein, C. R. (2015). *Which nudges do people like? A national survey*. Social Science Research Network. https://doi.org/10.2139/ssrn.2619899

van Oldenbeek, M., Winkler, T. J., Buhl-Wiggers, J., & Hardt, D. (2019, May 15). Nudging in blended learning: Evaluation of email-based progress feedback in a flipped-classroom information systems course. *Proceedings of the 27th European Conference on Information Systems, ECIS* (article 186). AIS eLibrary. https://aisel.aisnet.org/ecis2019_rp/186

Vasquez, H., Fuentes, A. A., Kypuros, J. A., & Azarbayejani, M. (2015, December 7). Early identification of at-risk students in a lower-level engineering gatekeeper course. In *2015 IEEE*

Frontiers in Education Conference, FIE (vol. 2015, pp. 1–9). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/FIE.2015.7344361

Weijers, R. J., de Koning, B. B., & Paas, F. (2020). Nudging in education: From theory towards guidelines for successful implementation. *European Journal of Psychology of Education, 36,* 883–902. https://doi.org/10.1007/s10212-020-00495-0

York, B., Loeb, S., & Doss, C. (2019). One step at a time: The effects of an early literacy text-messaging program for parents of preschoolers. *Journal of Human Resources*, *54*(3), 537–566. https://doi.org/10.3368/jhr.54.3.0517-8756R

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 23, Number 1

February – 2022

An Internet Articles Retrieval Agent Combined With Dynamic Associative Concept Maps to Implement Online Learning in an Artificial Intelligence Course

Yu-Ping Cheng[1], Shu-Chen Cheng[2], and Yueh-Min Huang[1]
[1]Department of Engineering Science, National Cheng Kung University, Taiwan; [2]Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Taiwan

Abstract

Online learning has been widely discussed in education research, and open educational resources have become an increasingly popular way to help learners acquire knowledge. However, these resources contain massive amounts of information, making it difficult for learners to identify Web articles that refer to computer science knowledge. This study developed an Internet articles retrieval agent combined with dynamic associative concept maps (DACMs). The system used text mining technology to analyze keywords to filter computer science articles. In previous research, concept maps were manually constructed; in this study, such maps can be automatically and dynamically generated in real time. In a case study of a fundamental course of artificial intelligence, this study designed two experiments to compare students' learning behaviors while using this system and the Google search engine. The results of the first experiment showed that the experimental group searched for more knowledge articles on computer science using this agent, compared to the control group using the Google search engine. The learning performance of the experimental group was significantly better than that of the control group, while the cognitive load of the experimental group was significantly lower than that of the control group. Furthermore, the results of the second experiment showed that the learning progress of students using the agent was significantly greater than that of students who used the Google search engine. This illustrates that the agent effectively filtered computer science articles, and DACMs helped students gain a deeper understanding of academic concepts and knowledge related to artificial intelligence.

Keywords: dynamic associative concept maps, text mining, online learning, intelligent agent

Introduction

In recent years, online learning has been widely discussed in education research (Abdullah & Mirza, 2020). Some studies indicated that online learning can effectively improve students' learning performance in the classroom (Dashtestani, 2020) and learning motivation (Chaiprasurt & Esichaikul, 2013; Li & Tsai, 2017). With the rapid development of information technology, more and more teachers provide online learning platforms and tools in the classroom so that students can conveniently use the Internet to implement online learning (Lowenthal et al., 2015; Tsai et al., 2018). Also, students can effectively obtain massive open resources and learning materials through online learning (Isaac et al., 2019). Compared with traditional lectures, books, and course resources, students can obtain more information and resources through online learning (Alshahrani et al., 2017). However, online open resources usually contain massive amounts of information, which may make it difficult for students to identify the connection between these resources and learning contents. Meanwhile, with the rapid development of artificial intelligence, students need to learn the latest knowledge and concepts very quickly.

Recent research has indicated that concept maps can effectively help online students understand new concepts while learning new knowledge (Farrokhnia et al., 2019). Concept maps have also been regarded as a useful tool for structural knowledge representation (Hwang et al., 2013; Hwang et al., 2011). In recent years, concept maps have been applied to different courses to improve students' learning performance. For example, Hwang, Zou, and Lin (2020) used a question-posing approach based on concept mapping to explore ubiquitous learning about plants in elementary natural science courses. Chiou et al. (2017) discussed various concept mapping techniques for senior accounting students. Chiou et al. (2015) used multimedia animation combined with multidimensional concept maps to discuss multimedia animation courses in universities; the results of these three studies showed that concept maps effectively promoted students' learning performance in the classroom. However, in previous studies, concept maps were usually constructed manually (Chularut & DeBacker, 2004; Marzano & Miranda, 2020; Sun & Chen, 2016); maps were not generated automatically and dynamically in real time, even as sources and amounts of information increased.

Current research has indicated that online learning combined with concept maps can effectively improve students' learning performance. For example, Fatawi et al. (2020) indicated that learning online can improve student learning outcomes and engagement through concept maps. Hwang, Chang, et al. (2020) indicated that a problem-posing strategy guided by concept maps and adopted in an online learning environment improved the learning performance of students. Compared with students with lower critical thinking, students with a higher level of critical thinking have a more obvious improvement in their learning performance. According to the survey, this study found no existing online learning platform that automatically filtered Web articles related to computer science and dynamically generated concept maps related to computer science in real time.

Our study sought to address the limitations of previous research; we developed an Internet articles retrieval agent combined with dynamic associative concept maps (DACMs). Students connected to this system for online learning through the Internet. This system used text mining technology to automatically filter computer science articles, and the Apriori algorithm automatically and dynamically generated associative concept maps in real time. In the context of a fundamental course in artificial intelligence, this study aimed to address the gaps in existing research, conquer the technical limitations of specific learning tools, reduce students' cognitive load, and improve their learning performance.

Literature Review

Online Learning

Many studies have indicated that online learning is a way to gain a learning experience through the use of certain technologies (Benson, 2002; Carliner, 2004; Conrad, 2002). Hwang, Wang, and Lai (2020) indicated that online learning can be independent of time and place, allowing students to watch and work with learning materials or multimedia videos ubiquitously. These online resources include text, pictures, videos, databases, and so on (Elbaum et al., 2002). Although many studies have confirmed the convenience of online learning, Doo et al. (2020) indicated that it requires effective learning methods to improve the quality of learning in the online environment. Therefore, this study proposed the use of DACMs with online learning activities as students searched for and read computer science articles autonomously.

Concept Maps

A concept map is a graphical tool commonly used to organize or express knowledge (Novak & Cañas, 2006). Students connected and integrated the concepts of new knowledge with their own knowledge through the use of concept maps (Chiou et al., 2015). Previous research has indicated that concept maps can effectively promote meaningful learning (Farrokhnia et al., 2019). In recent years, many researchers have confirmed the effectiveness of concept maps in education. For example, use of concept maps have improved students' learning performance (Hwang, Zou, & Lin, 2020), learning motivation (Hsu, 2019), problem solving (Whitelock-Wainwright et al., 2020), and critical thinking (Yue et al., 2017). In addition, Sun and Chen (2016) proposed dynamic concept maps with IRS to investigate the learning effects for students in anti-phishing education, and this map was gradually shown by the teachers' instructional explanations. This study showed that the use of dynamic concept maps with IRS by students can effectively improve their learning outcomes. Marzano and Miranda (2020) proposed a dynamap remediation approach that allowed users to specify nodes, relationships, and related content as they created new dynamic concept maps. However, these maps were manually constructed by the instructor in advance. For now, new concept maps cannot be automatically generated as open resources and information grow rapidly.

This study established an Internet articles retrieval agent combined with DACMs to realize the dynamic generation of associative concept maps in real time. Because this system was built on a server, DACMs were generated automatically and dynamically in real time whenever information was updated. These DACMs effectively conquered the limitation that meant instructors needed to manually construct concept maps in advance. This study designed two experiments to verify the effectiveness of the system in implementing online learning activities in a fundamental course on artificial intelligence, and was guided by the following research questions:

1. In Experiment 1, is the learning performance of the experimental group using the Internet articles retrieval agent combined with DACMs higher than that of the control group using the Google search engine?

2. In Experiment 1, is the cognitive load of the experimental group using the Internet articles retrieval agent combined with DACMs lower than that of the control group using the Google search engine?

3. In Experiment 2, is the learning progress of the same group using the Internet articles retrieval agent combined with DACMs higher than those who use the Google search engine?

Method

In this study, an Internet articles retrieval agent combined with DACMs was used for online learning activities in a basic course on artificial intelligence in order to explore students' learning performance, cognitive load, and learning progress. Two experiments were designed to compare the learning behaviors of students using this system with those using the Google search engine.

Experiment 1

Participants

A total of 75 college students in the department of computer science and information engineering at a university in Taiwan participated in this experiment. The ages of the participants ranged from 20 to 21 years. None of the students had previous experience with an online learning activity involving the use of an Internet articles retrieval agent combined with DACMs.

Research Procedure and Artificial Intelligence Learning Activity

Figure 1 shows the research procedure for Experiment 1, based on a quasi-experimental design method. Of the 75 students who participated, 38 students were assigned to the experimental group, and 37 students were assigned to the control group. The experimental group used the Internet articles retrieval agent combined with DACMs in specific online learning activities, and the control group used the Google search engine. The same teacher instructed the online learning activities for both groups. The experimental time for each group was the duration of one class period (50 minutes).

Before the experiment began, the teacher introduced the experimental procedure to the students. When the experiment started, all students in both groups were given a pre-test to measure their knowledge of concepts related to a basic artificial intelligence course. Next, the teacher introduced the learning tools to be used for implementing online learning activities, namely the Internet articles retrieval agent combined with DACMs or the Google search engine. Each group then used the specifically designated learning tools to implement their learning tasks. Finally, all students in both groups were given a post-test and cognitive load questionnaire to explore whether there were significant between-group differences in terms of learning performance and cognitive load after using the designated learning tools.

Figure 1

Research Procedure for Experiment 1

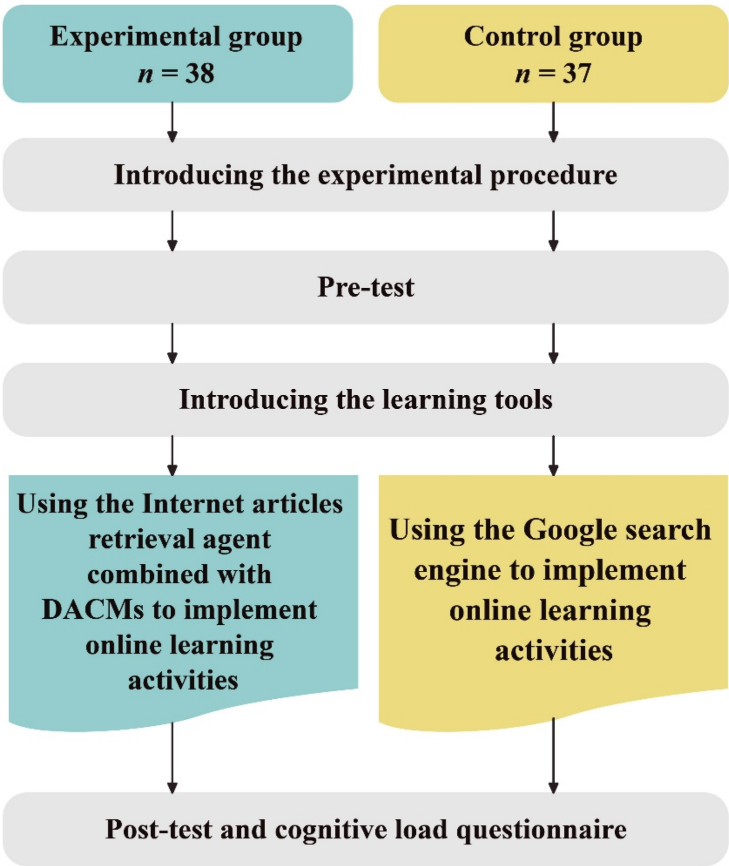


Figure 2 shows the learning conditions for the experimental group who used the Internet articles retrieval agent combined with DACMs. In this learning task, the teacher assigned a keyword related to artificial intelligence. The experimental group searched for articles with this keyword using the proposed system, recorded the titles and the corresponding URLs, and read the articles they found. In addition to searching for computer science articles through this system, the experimental group visualized the relevance between each keyword and related keywords through DACMs.

Figure 2

Learning Conditions for the Experimental Group

The learning conditions for the control group who used the Google search engine to complete the same learning task performed by the experimental group. This ensured consistency between the two groups in terms of learning difficulty. The keywords that the control group used in their searches were the same as the experimental group.

Experiment 1 was designed to explore the learning performance and cognitive load of the two groups in terms of online learning activities. The teacher assigned the keywords that students had not yet learned. In order to ensure that the analytical results of the post-test were correct and reliable, the pre-test answers were not given to the students. The students autonomously searched for articles associated with the keywords and read the computer science articles they found.

Experiment 2

Participants

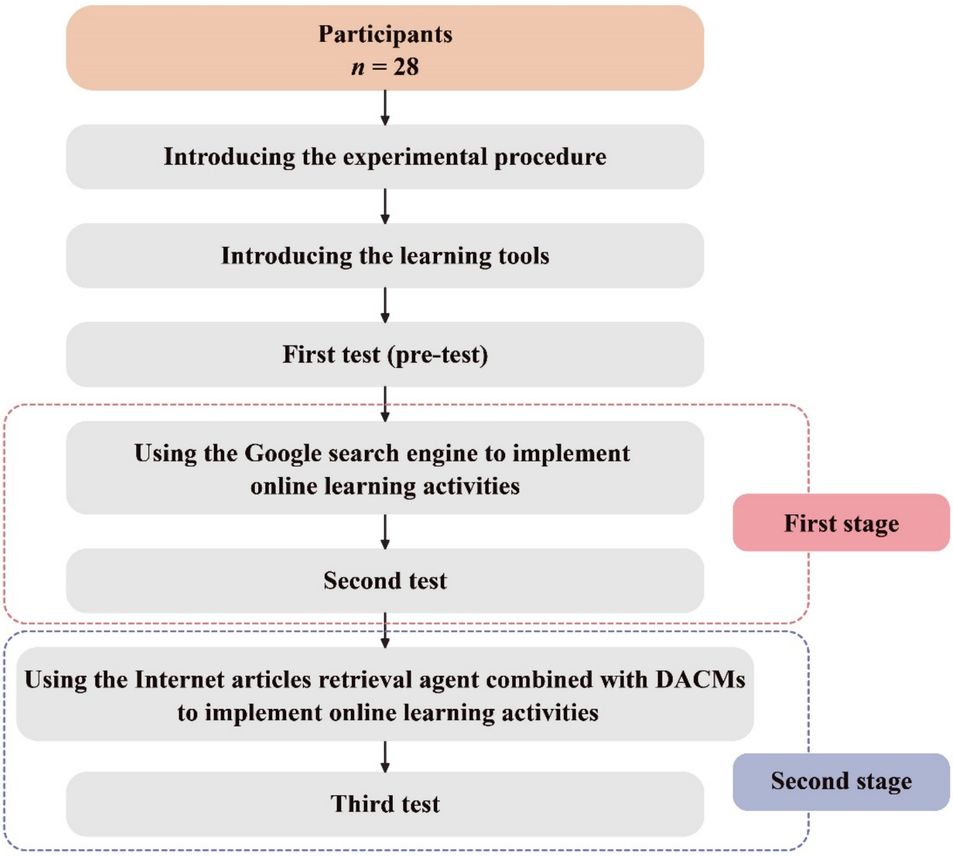
A total of 28 students in the department of computer science and information engineering at a five-year junior college at a university in Taiwan volunteered to participate in Experiment 2. Participants, who were selected based on availability and willingness to participate, ranged in age from 16 to 17 years. All students used the Internet articles retrieval agent combined with DACMs and the Google search engine in two stages to implement the online learning activities. Before the start of the online learning activity, none of the students had experience with the Internet articles retrieval agent combined with DACMs.

Research Procedure and Artificial Intelligence Learning Activity

Figure 3 shows the research procedure for Experiment 2 which was implemented in two stages for the duration of one class period. First, the teacher explained the experimental procedure and then introduced the two learning tools to the students. Next, all students took the first test (i.e., pre-test) to determine their level of prior knowledge before using any learning tools. In the first stage of the experiment, the students used the Google search engine for the assigned learning tasks. After completing the learning task, the students took the second test to assess their learning performance after using the Google search engine. In the second stage of the experiment, the students used the Internet articles retrieval agent combined with DACMs to complete the assigned learning tasks. After completing the learning task, the students took the third test to determine their learning performance after they used the proposed system. All the students used two learning tools, one in each of the two phases; three tests conducted in two stages explored whether there were significant differences in learning progress depending on the tools used.

Figure 3

Research Procedure for Experiment 2



The teacher assigned a different keyword in the second stage with the same difficulty as the keyword in the first stage. This prevented students from repeating what they already had learned. The learning tasks implemented in each stage of Experiment 2 was the same as that in Experiment 1.

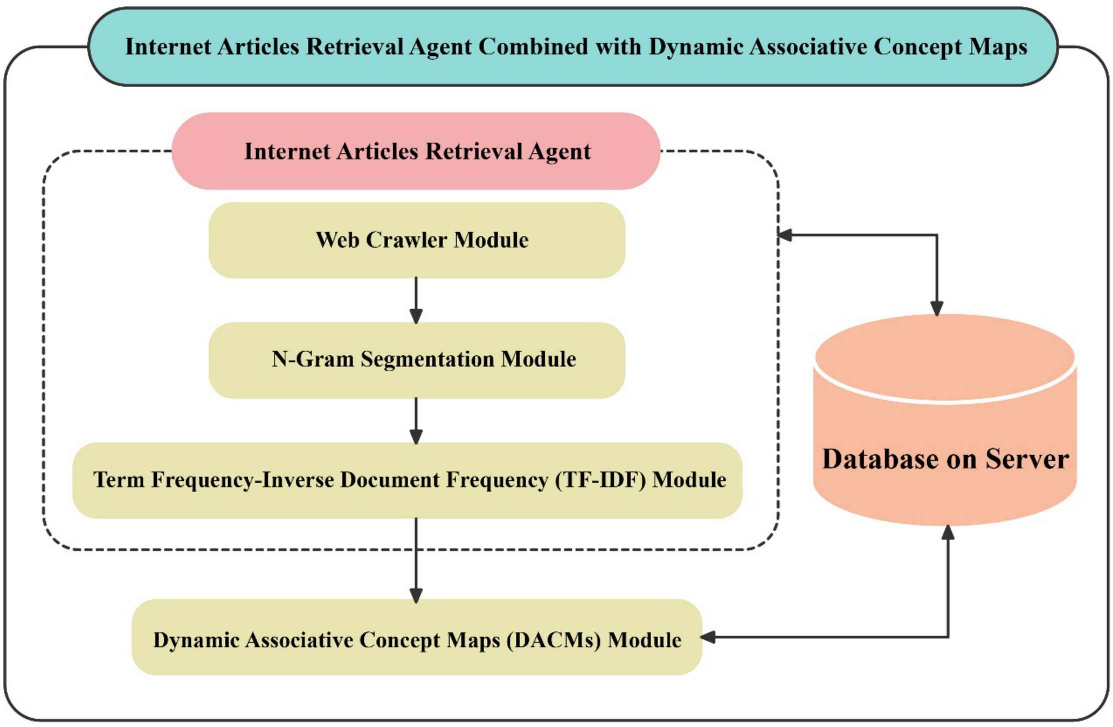
In Experiment 2, this study determined whether there were significant differences in learning progress when the students used two different learning tools to carry out online learning activities. When the students used the Google search engine, their learning progress was based on the difference in scores between the first test and the second. When the students used the Internet articles retrieval agent combined with DACMs, their corresponding learning progress was based on the difference in the scores between the second test and the third. In order to ensure that the analytical results were correct and reliable, the test answers were not provided to the students. The students autonomously searched for articles associated with the keywords and read the computer science articles they found.

System Architecture and Functionality

This study used text mining technology to develop an Internet articles retrieval agent combined with DACMs to improve students' learning performance in a fundamental course on artificial intelligence. The system was built on the server, so students learned online through the Internet at any time. The system collected a total of 35,907 computer science articles that contained 1,751 keywords. Text mining technology automatically filtered the articles and provided main keywords from the articles, and the Apriori algorithm automatically and dynamically generated associative concept maps in real time. The unique advantage of this system was that it allowed students to quickly search through computer science articles for the latest information on artificial intelligence. The use of DACMs strengthened the keywords and concepts related to AI. Figure 4 illustrates the system architecture for this study.

Figure 4

System Architecture Diagram



This system used the Web crawler module to retrieve an enormous number of online articles. The N-gram segmentation module pre-processed the Web articles and automatically filtered to select those that dealt with computer science. Then, the term frequency-inverse document frequency module automatically calculated the text weighting of each computer science article. Finally, each computer science article with main keywords were stored in the database on the server. In addition, the DACMs module used the Apriori algorithm to automatically determine the strength of association between keywords, so as to dynamically generate the associative concept map in real time.

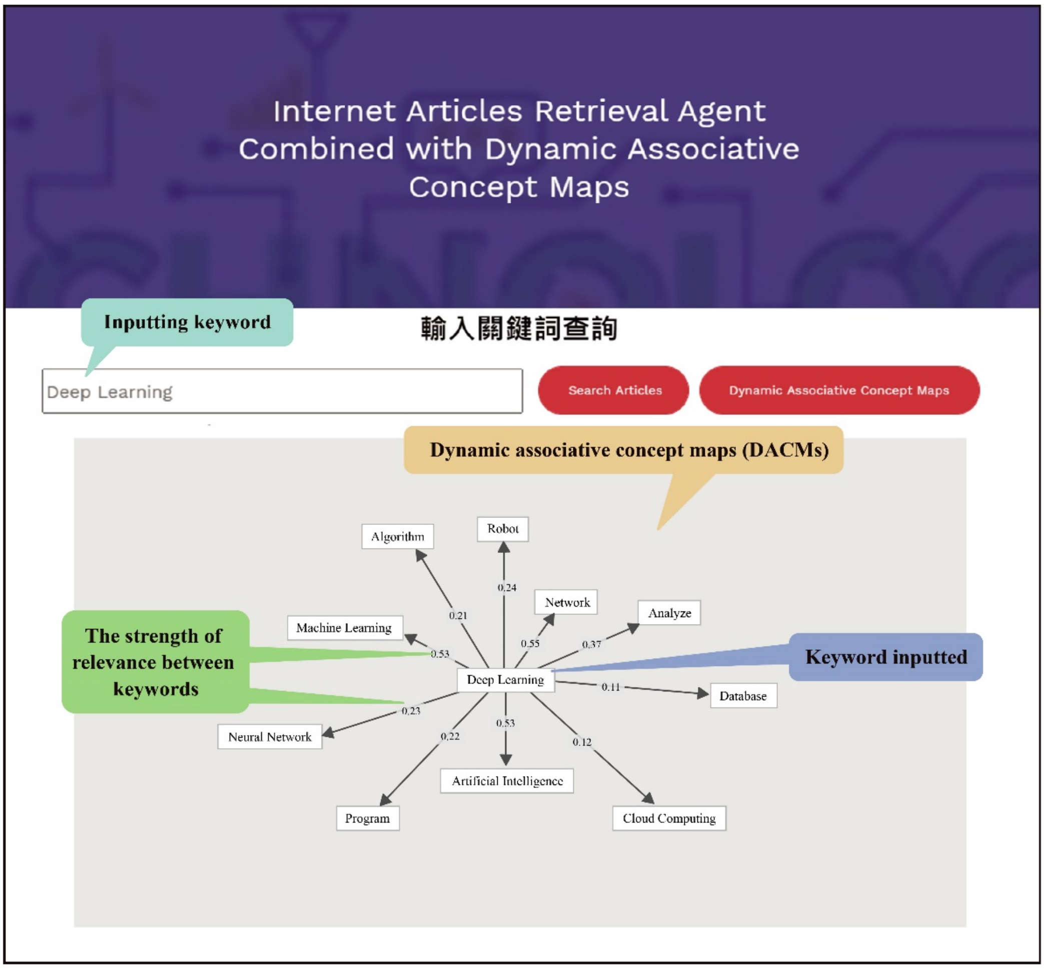
Dynamic Associative Concept Maps (DACMs)

In previous studies, concept maps were not generated automatically; teachers constructed concept maps manually for specific courses or topics. Artificial intelligence is developing at a rapid rate, and the source of knowledge is no longer limited to textbooks. In order to enable students to obtain the latest information and concepts any time and any place, this study used DACMs based on the Apriori algorithm and text mining technology, so that concept maps were generated dynamically in real time. Through this function, the proposed system solved the limitations of previous research in which concept maps were constructed manually. In addition, students also accessed and worked with dynamic associative concept maps with the latest AI information in real time. Compared to common concept maps, DACMs automatically and dynamically generated the strength of the correlations between keywords in real time. In addition to quickly understanding the relevance of various keywords through the DACMs, students also learned highly-related keywords to further enhance their learning performance.

Figure 5 is an example of a DACM used in the proposed system. Students entered keywords, and the system generated dynamic associative concept maps in real time. The DACMs calculated the relevance between each keyword, and automatically connected and displayed the strength of relevance between keywords in the arrows and the association rule. In addition, students used the DACMs to learn new keywords, which promoted a deeper understanding of the keywords and enhanced their learning of a new concept.

Figure 5

Sample Dynamic Associative Concept Map (DACM) for the Keywords Deep Learning



Internet Articles Retrieval Agent

The Internet articles retrieval agent used text mining technology to automatically filter computer science articles so that students can avoid reading articles unrelated to computer science. Therefore, this agent provided students with an efficient way to search for computer science articles any time and any place, and also provided keywords for each computer science article. Figure 6 shows the operation interface for the Internet articles retrieval agent used to search for computer science articles. The system provided the title and link to a computer science article as well as main keywords for the article. Before they read each article, the students reviewed the main keywords and determined whether the article was suitable. Therefore, the computer science articles and main keywords provided by this system helped students gain a deeper understanding of new information related to artificial intelligence.

Figure 6

Operation Interface for the Internet Articles Retrieval Agent

Data Collection

Quantitative data were collected from the two experiments. In Experiment 1, data were collected from the pre-test, post-test, and cognitive load questionnaires filled out by 75 students, and in Experiment 2, from the first, second, and third test results collected from 28 students. The following provides details on the data collected.

- Test content: 10 keywords related to computer science were selected. The students indicated which keywords were highly relevant to artificial intelligence. Each question was worth 10 points, with a full score of 100 points.

- Cognitive load questionnaire: This study was modified from the measure developed by Hwang et al. (2013) and measured the cognitive load for the students in Experiment 1. This questionnaire comprised two dimensions with a total of eight items. Five items measured mental load and three items measured mental effort. All responses used a five-point Likert scale.

Data Analysis

SPSS statistical analysis software was used to analyze data from the two experiments.

- Experiment 1: Analysis of covariance (ANCOVA) and independent sample *t* test were used to analyze the learning performance and cognitive load of 75 students. In addition, Cronbach's alpha was used to check the internal consistency of the cognitive load questionnaire. The Cronbach's alpha for mental load was 0.82, and 0.85 for mental effort. These results indicated that the cognitive load questionnaire had good reliability.
- Experiment 2: Paired sample *t* tests were used to analyze learning progress for 28 students.

Results

Experiment 1

Learning Performance

Before adopting ANCOVA to analyze students' learning performance in Experiment 1, this study conducted a homogeneity regression to ensure there was no interaction between the independent variables and covariates. For the independent variables, the experimental group used the Internet articles retrieval agent combined with DACMs to implement online learning activities, and the control group used the Google search engine. The covariates included the pre-test taken by the two groups of students. These results showed that there were no significant differences in the interaction between the independent variables and the covariates ($F = 2.81$, $p > .05$).

Table 1 shows ANCOVA results for the post-test. The adjusted mean and standard error of the post-test for the experimental group were 53.09 and 2.48, respectively, and the adjusted mean and standard error of the post-test of the control group were 41.42 and 2.52, respectively. The results of ANCOVA showed that by excluding the influence of the pre-test, the post-test of the two groups reached a significant difference ($F = 10.9$, $p < .01$). This means that the students in the experimental group using the Internet articles retrieval agent combined with DACMs exhibited significantly better learning performance as compared to the control group using the Google search engine.

Table 1

ANCOVA Results for Learning Performance in Experiment 1

Group	n	Mean	SD	Adjusted mean	Std. error	F	η^2
Experimental	38	52.89	16.43	53.09	2.48	10.9**	0.13
Control	37	41.62	15.37	41.42	2.52		

Note. **$p < .01$.

In addition, we counted the number of computer science articles searched for by the two groups during the learning tasks. On average, each student in the experimental group found 9.24 computer science articles, whereas each student in the control group found 3.41 computer science articles. Clearly, the main keywords in each computer science article provided by the proposed system effectively helped students to determine the key points of the article from the start. This not only helped the students search for and read more computer science articles but also sped up their judgments as to whether the article was related to the search keywords. In addition, the experimental group significantly improved in terms of learning performance after using the proposed system. This suggests that DACMs promoted student understanding of the strength of the associations between keywords, and in turn, effectively enhanced their acquisition of professional information and concepts related to artificial intelligence.

Cognitive Load

An independent sample *t* test was used to analyze the cognitive load of the students during Experiment 1; Table 2 shows the results. The mean and standard deviations of cognitive load for the experimental group were 1.63 and 0.58, respectively, and the mean and standard deviation of cognitive load for the control group were 2.01 and 0.6, respectively. The independent sample *t* test results showed that the cognitive load in the two groups reached a significant difference ($t = -2.79$, $p < .01$). This means that the students in the experimental group had a significantly lower cognitive load than did the students in the control group using the Google search engine. In other words, when the students in the experimental group used the proposed system to acquire the professional information and concepts related to artificial intelligence, they did not experience an increase in cognitive load that influenced their learning process.

Table 2

Independent Sample t Test Results for Cognitive Load in Experiment 1

Group	*n*	Mean	*SD*	*t*
Experimental	38	1.63	0.58	-2.79**
Control	37	2.01	0.6	

Note. $^{**}p < .01$.

This study also explored two dimensions of cognitive load. Table 3 shows the independent sample *t* test results for mental load and mental effort. The means of mental load and mental effort for the experimental group were 1.68 and 1.54, respectively, and the mean of mental load and mental effort for the control group were 2.09 and 1.87, respectively. The independent sample *t* test results showed significant between-group differences between mental load ($t = -2.9$, $p < .01$) and mental effort ($t = -2.06$, $p < .05$). This indicated that when the experimental group used the proposed system to complete the learning task, their mental load and mental effort were significantly lower than was the case for the control group using the Google search engine to complete the learning task. In other words, the Internet articles retrieval agent combined with DACMs effectively assisted the students in the experimental group to search for correct computer science articles related to the learning task, so they successfully completed it.

Table 3

Independent Sample t Test Results for Mental Load and Mental Effort

Dimension	Group	*n*	Mean	*SD*	*t*
Mental load	Experimental	38	1.68	0.59	-2.9**
	Control	37	2.09	0.62	
Mental effort	Experimental	38	1.54	0.64	-2.06*
	Control	37	1.87	0.75	

Note. $^{*}p < .05$, $^{**}p < .01$.

In summary, there were significant between-group differences in both the analysis of cognitive load and in the analysis of mental load and mental effort. The Internet articles retrieval agent combined with

DACMs effectively reduced the cognitive load of students in the experimental group. At the same time, the students were able to correctly search for computer science articles related to the learning task. Therefore, it is inferred that the proposed system effectively helped these students read about and acquire professional knowledge and concepts related to artificial intelligence, while at the same time it reduced the cognitive load in their learning process.

Experiment 2

A paired sample *t* test was used to analyze improvements in learning progress in Experiment 2. Table 4 provides the paired samples statistics for the three tests. The mean and standard deviation for the first test were 25.71 and 19.89, respectively. The mean and standard deviation for the second test were 34.64 and 22.69, respectively, and for the third test, 52.5 and 15.06, respectively.

Table 5 shows the results of the paired sample *t* test for learning progress. The average progress and standard deviation from learning performance for those who used the Google search engine were 8.93 and 13.43, respectively. The average progress and standard deviation from learning performance for those who used the Internet articles retrieval agent combined with DACMs were 17.86 and 14.49, respectively. The paired sample *t* test results indicated that the improvement in learning progress in the two stages reached a significant difference (t = -2.18, p < .05). This means that compared with the average progress of students using the Google search engine (25.71 in the first test, 34.64 in the second test, average progress of 8.93), students using the proposed system had greater average progress (34.64 in the second test, 52.5 in the third test, progress average of 17.86). Therefore, the improvement in learning progress for the students using the Internet articles retrieval agent combined with DACMs was significantly better than for those using the Google search engine. The proposed system effectively helped the students search for and read computer science articles related to artificial intelligence, and in turn enhanced their professional knowledge and understanding of concepts related to it.

Table 4

Paired Samples Statistics for the Three Tests in Experiment 2

Test	*n*	Mean	*SD*
First	28	25.71	19.89
Second	28	34.64	22.69
Third	28	52.5	15.06

Table 5

The Paired Sample t Test Results for Using the Different Learning Tools

Learning tool	*n*	Progress average	*SD*	*t*
Google search engine	28	8.93	13.43	-2.18*
Internet articles retrieval agent combined with DACMs	28	17.86	14.49	

Note. *p < .05.

We counted the number of computer science articles found by students using the different learning tools in the two stages. Each student using the Internet articles retrieval agent combined with DACMs found an average of 3.0 computer science articles; each student using the Google search engine found an average of 2.57 computer science articles. In Experiment 2, the number of computer science articles found by students using the proposed system and Google search engine were similar.

Discussion and Conclusions

In this study, an Internet articles retrieval agent combined with DACMs was used to improve students' learning performance in a fundamental artificial intelligence course while reducing their cognitive load when learning new information and concepts. In addition, the DACMs proposed in this study were based on the Apriori algorithm and text mining technology, which automatically and dynamically generated a concept map in real time and mitigated the limitations related to manual construction of concept maps found in previous studies. DACMs not only helped students understand the strength of relevance between keywords or concepts but also automatically and dynamically generated related concept maps with the latest information in real time; students used this system to learn online at any time. Two experiments were conducted to compare the learning behavior of students using the proposed system with the Google search engine.

Experiment 1

The results for Experiment 1 showed significant improvements in learning performance in the experimental group as compared to the control group. The results of this study confirmed that the Internet articles retrieval agent combined with DACMs significantly improved students' learning performance. It is thus inferred that DACMs effectively provided students with knowledge about the strength of the correlation between various keywords related to artificial intelligence and so helped students understand the latest information or concepts. At the same time, the students were able to use DACMs to recognize keywords that had not been learned previously, as well as to use the Internet articles retrieval agent to search for computer science articles. Compared to the control group using the Google search engine, the experimental group was able to search for more computer science articles; the system promoted students' learning of information and concepts related to artificial intelligence while they consolidated the latest information.

In addition, the cognitive load results showed that the experimental group's cognitive load was significantly lower than that of the control group. At the same time, the results showed that the mental load and mental effort exerted by the experimental group were significantly lower than that of the control group. The results of this study confirmed that the use of the proposed system to acquire professional knowledge and understand concepts related to artificial intelligence did not cause excessive cognitive load during the learning process. It is thus inferred that the system effectively enabled the students to search for computer science articles related to the learning tasks, and the DACMs made it possible for them to visualize the strength of the correlations between keywords. The students were able to easily understand and learn artificial intelligence keywords through this system, and they were also able to search for and read artificial intelligence computer science articles and successfully complete the learning tasks.

Experiment 2

The results of Experiment 2 indicated that the students using the Internet articles retrieval agent combined with DACMs had significantly greater improvements in learning progress compared to the control group using the Google search engine. The results of this study confirmed that the use of the

Internet articles retrieval agent combined with DACMs enhanced the students' knowledge acquisition and understanding of relevant concepts to a greater degree than did the use of the Google search engine. In addition, the teacher assigned different keywords in the two stages to compare the learning progress of the students using different learning tools. Although the number of computer science articles found by students using the proposed system was not very different from that found by the students using the Google search engine, it is worth noting that the quality of computer science articles the students searched for might affect the differences in learning effectiveness. The students using the proposed system to search for and read computer science articles exhibited significantly greater learning progress as compared to the students using the Google search engine. Therefore, Experiment 2 showed that, compared to the Google search engine, the proposed system provided artificial intelligence-related computer science articles more effectively, and made it possible for students to gain deeper insights into the latest information and concepts in the same amount of experimental time.

In summary, the results of this study can be used by educators to verify the importance of students' acquiring novel information and concepts in open educational resources. The system proposed in this study is open and free. Students can use this system to achieve online learning and distance education through the Internet. Furthermore, this study suggests that an Internet articles retrieval agent combined with DACMs can serve as a superior auxiliary learning tool for students using open educational resources, and students can acquire novel information and concepts using this system.

References

Abdullah, N. A., & Mirza, M. S. (2020). Evaluating pre-service teaching practice for online and distance education students in Pakistan: Evaluation of teaching practice. *The International Review of Research in Open and Distributed Learning, 21*(2), 81–97. https://doi.org/10.19173/irrodl.v21i2.4606

Alshahrani, S., Ahmed, E., & Ward, R. (2017). The influence of online resources on student-lecturer relationship in higher education: A comparison study. *Journal of Computers in Education, 4*(2), 87–106. https://doi.org/10.1007/s40692-017-0083-8

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

Carliner, S. (2004). *An overview of online learning* (2nd ed.). Amherst, MA: Human Resource Development Press.

Chaiprasurt, C., & Esichaikul, V. (2013). Enhancing motivation in online courses with mobile communication tool support: A comparative study. *International Review of Research in Open and Distributed Learning, 14*(3), 377–401. https://doi.org/10.19173/irrodl.v14i3.1416

Chiou, C. C., Lee, L. T., Tien, L. C., & Wang, Y. M. (2017). Analyzing the effects of various concept mapping techniques on learning achievement under different learning styles. *EURASIA Journal of Mathematics, Science and Technology Education, 13*(7), 3687–3708. https://doi.org/10.12973/eurasia.2017.00753a

Chiou, C. C., Tien, L. C., & Lee, L. T. (2015). Effects on learning of multimedia animation combined with multidimensional concept maps. *Computers & Education, 80*, 211–223. https://doi.org/10.1016/j.compedu.2014.09.002

Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29*(3), 248–263. https://doi.org/10.1016/j.cedpsych.2003.09.001

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

Dashtestani, R. (2020). Online courses in the higher education system of Iran: A stakeholder-based investigation of pre-service teachers' acceptance, learning achievement, and satisfaction. *The International Review of Research in Open and Distributed Learning, 21*(4), 117–142. https://doi.org/10.19173/irrodl.v21i4.4873

Doo, M. Y., Tang, Y., Bonk, C. J., & Zhu, M. (2020). MOOC instructor motivation and career development. *Distance Education, 41*(1), 26–47. https://doi.org/10.1080/01587919.2020.1724770

Elbaum, B., McIntyre, C., & Smith, A. (2002). *Essential elements: Prepare, design, and teach your online course*. Atwood Publishing.

Farrokhnia, M., Pijeira-Díaz, H. J., Noroozi, O., & Hatami, J. (2019). Computer-supported collaborative concept mapping: The effects of different instructional designs on conceptual understanding and knowledge co-construction. *Computers & Education, 142*, 103640. https://doi.org/10.1016/j.compedu.2019.103640

Fatawi, I., Degeng, I. N. S., Setyosari, P., Ulfa, S., & Hirashima, T. (2020). Effect of online-based concept map on student engagement and learning uutcome. *International Journal of Distance Education Technologies, 18*(3), 42–56. https://doi.org/10.4018/IJDET.2020070103

Hsu, T. C. (2019). Using a concept mapping strategy to improve the motivation of EFL students in Google Hangouts peer-tutoring sessions with native speakers. *Interactive Learning Environments, 27*(2), 272–285. https://doi.org/10.1080/10494820.2018.1463268

Hwang, G. J., Chang, S. C., Song, Y., & Hsieh, M. C. (2020). Powering up flipped learning: An online learning environment with a concept map-guided problem-posing strategy. *Journal of Computer Assisted Learning, 37*(2), 429–445. https://doi.org/10.1111/jcal.12499

Hwang, G. J., Shi, Y. R., & Chu, H. C. (2011). A concept map approach to developing collaborative Mind Tools for context-aware ubiquitous learning. *British Journal of Educational Technology, 42*(5), 778–789. https://doi.org/10.1111/j.1467-8535.2010.01102.x

Hwang, G. J., Wang, S. Y., & Lai, C. L. (2020). Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics. *Computers & Education, 160*, 104031. https://doi.org/10.1016/j.compedu.2020.104031

Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education, 69*, 121–130. https://doi.org/10.1016/j.compedu.2013.07.008

Hwang, G. J., Zou, D., & Lin, J. (2020). Effects of a multi-level concept mapping-based question-posing approach on students' ubiquitous learning performance and perceptions. *Computers & Education, 149*, 103815. https://doi.org/10.1016/j.compedu.2020.103815

Isaac, O., Aldholay, A., Abdullah, Z., & Ramayah, T. (2019). Online learning usage within Yemeni higher education: The role of compatibility and task-technology fit as mediating variables in the IS success model. *Computers & Education, 136*, 113–129. https://doi.org/10.1016/j.compedu.2019.02.012

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

Lowenthal, P., Bauer, C., & Chen, K. Z. (2015). Student perceptions of online learning: An analysis of online course evaluations. *American Journal of Distance Education, 29*(2), 85–97. https://doi.org/10.1080/08923647.2015.1023621

Marzano, A., & Miranda, S. (2020). The dynamap remediation approach (DMRA) in online learning environments. *Computers & Education*, 104079.

https://doi.org/10.1016/j.compedu.2020.104079

Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them. *Florida Institute for Human and Machine Cognition, 1*(1), 1–31.

Sun, J. C. Y., & Chen, A. Y. Z. (2016). Effects of integrating dynamic concept maps with interactive response system on elementary school students' motivation and learning outcome: The case of anti-phishing education. *Computers & Education, 102*, 117–127. https://doi.org/10.1016/j.compedu.2016.08.002

Tsai, Y. H., Lin, C. H., Hong, J. C., & Tai, K. H. (2018). The effects of metacognition on online learning interest and continuance to learn with MOOCs. *Computers & Education, 121*, 18–29. https://doi.org/10.1016/j.compedu.2018.02.011

Whitelock-Wainwright, A., Laan, N., Wen, D., & Gašević, D. (2020). Exploring student information problem solving behaviour using fine-grained concept map and search tool data. *Computers & Education, 145*, 103731. https://doi.org/10.1016/j.compedu.2019.103731

Yue, M., Zhang, M., Zhang, C., & Jin, C. (2017). The effectiveness of concept mapping on development of critical thinking in nursing education: A systematic review and meta-analysis. *Nurse Education Today, 52*, 87–94. https://doi.org/10.1016/j.nedt.2017.02.018

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 23, Number 1

February – 2022

The Effects on Secondary School Students of Applying Experiential Learning to the Conversational AI Learning Curriculum

Ting-Chia Hsu[1], Hal Abelson[2], and Jessica Van Brummelen[2*]
[1]Department of Technology Application and Human Resource Development, National Taiwan Normal University;
[2]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA

Abstract

The purpose of this study was to design a curriculum of artificial intelligence (AI) application for secondary schools. The learning objective of the curriculum was to allow students to learn the application of conversational AI on a block-based programming platform. Moreover, the empirical study actually implemented the curriculum in the formal learning of a secondary school for a period of six weeks. The study evaluated the learning performance of students who were taught with the cycle of experiential learning in one class, while also evaluating the learning performance of students who were taught with the conventional instruction, which was called the cycle of doing projects. Two factors, learning approach and gender, were taken into account. The results showed that females' learning effectiveness was significantly better than that of males regardless of whether they used experiential learning or the conventional projects approach. Most of the males tended to be distracted from the conversational AI curriculum because they misbehaved during the conversational AI process. In particular, in their performance using the Voice User Interface with the conventional learning approach, the females outperformed the males significantly. The results of two-way ANCOVA revealed a significant interaction between gender and learning approach on computational thinking concepts. Females with the conventional learning approach of doing projects had the best computational thinking concepts in comparison with the other groups.

Keywords: gender studies, conversational AI application, experiential learning, block-based programming

Introduction

In the technology era, from understanding complex terminology, syntax, and error messages, to learning about functions, iterations, and new algorithms, some students, even at the university level, have difficulty learning to program (Piwek & Savage, 2020). Because of this, many researchers have investigated innovative and useful approaches for teaching and learning computer programming. For example, researchers have proposed an experiential learning cycle from project-based learning for learning computer science (Pucher & Lehner, 2011). These methods involved concrete experience, the application of acquired knowledge, the contextualization of projects in the real world, and hands-on implementation, which are highly relevant to developing computer programs (Efstratia, 2014; Sendall et al., 2019).

With the fast-paced, continual development of computer science, including huge gains in artificial intelligence (AI) and machine learning, the application of AI has become popular in our daily lives due to the high-speed development of hardware (Hsu et al., 2021). One rapidly-growing subfield includes conversational AI, which is the ability of machines to converse with humans, including voice-based technologies such as Amazon's Alexa. The goal of the current study was therefore to investigate the effectiveness of using the cycle of experiential learning and the cycle of doing projects in a conversational AI curriculum. Specifically, this research investigated the two different teaching approaches—the cycle of experiential learning and the conventional cycle of doing projects—with a visual programming interface for conversational AI applications using the MIT App Inventor (Van Brummelen, 2019). The conversational AI curriculum we developed allowed young students to connect the application of audio interaction with the Internet of things (IoT) or simulative interaction in the block-based programming environment. This innovative, applied AI curriculum was designed to be implemented in junior high schools.

For novices and young students, there is evidence that visual programming, which is also termed block-based programming, is more effective in teaching programming than is conventional command-line programming with complex syntax (Cetin, 2016). In this study, visual programming tools referred to block-based programming tools such as MIT App Inventor or Scratch. In comparison with conventional text-based programming, such visual programming tools have been helpful for novices to fully focus on learning to solve problems as well as understand the logic and framework of the overall program, rather than attend to specific semantics or syntax (Grover & Pea, 2013; Hsu et al., 2018; Lye & Koh, 2014).

Literature Review

In conventional programming, programs are written with strict syntax, which can be difficult for general populations to learn, especially non-native English speakers, since a program cannot run successfully it has even minor spelling errors. On the other hand, if students use block-based programming to build the program, these errors will not occur. Block-based programming emphasizes recognition over recall; code-blocks are readily available in the visual interface. Furthermore, the blocks are categorized according to their function or logic. Students only need to concentrate on using appropriate blocks to complete the work they want to do or to create the effect they desire, rather than memorize syntax or particular keywords of the programming language. Moreover, the shape and color of the blocks provide the students with scaffolding to emphasize which blocks can be linked together and how code can (or cannot) be developed. During this process of visual code development, students learn the concepts of composing programs and that different blocks have various functions or properties. With block-based programming, students usually need only drag and connect the blocks, reducing the cognitive load and allowing students to focus on the logic and structures involved in programming rather than the syntax of writing programs (Kelleher & Pausch, 2005). Block-based programming provides students with media-rich learning environments, allowing them to connect with various personal interests (Brennan & Resnick, 2012). Chiu (2020) discovered that learners were very positive about the creation of applications (apps) by visual programming and project development, and recommended that novice programmers create apps with block-based programming. Finally, when students used a visual programming tool to write a program, they tended to focus on solving problems. Researchers have indicated that visual programming tools have a positive impact on programming self-efficacy and decrease student frustration (Yukselturk & Altiok, 2017).

It is especially important to reduce learning frustration for those who are underrepresented in computer science, as they face additional challenges when they first enter the field. Furthermore, it is important to increase their participation in computer science, as underrepresented groups provide unique perspectives and diverse, innovative solutions. In this paper, we investigated the effectiveness of different learning techniques by gender, since historically, females have been underrepresented in computer science, and the relative number of females entering the field has significantly decreased over the past 30 years (Weston et al., 2019). By determining and using the most effective pedagogical techniques for computer science by gender, more females may enter the field, and the gender gap may close.

A previous study has shown that gender impacted the ease of use and intention to use block-based programming (Cheng, 2019). Nonetheless, very little is known about the effect of gender on learning computational thinking skills in primary and secondary education (Kalelioğlu, 2015). Due to the shortage of females participating in science, technology, engineering and mathematics (STEM) domains in comparison with the number of males, many countries have recently encouraged females to participate in those domains. However, researchers have indicated that the participation rate of females is still lower than that of males in the computer science domain (Cheryan et al., 2017). The difference in male and female interest in computer science likely originates from females having less experience learning computer science during childhood (Adya & Kaiser, 2005).

Information processing theory research has also indicated that different genders have different perceptions and processing modes in the brain (Meyers-Levy, 1986). Males tended to rely on the right brain to process and select the input information from outside. Thus, they often paid attention to visual information or contextual signals, while ignoring the details of processing methods (Meyers-Levy, 1989). Conversely, females tended to prefer using their left brain to accept and analyze the input information in detail, often resulting in higher stress levels. Moreover, females tended to relate, collaborate, and share information with others (Putrevu, 2001). Different genders have different information processing procedures in the brain, and tended to filter and accept different types of input from the same information (Martin et al., 2002). Accordingly, it is worth exploring the effect of gender on new curricula such as the conversational AI curriculum with MIT App Inventor.

A previous study has shown there was no significant difference between genders in students' performance when programming using code.org, although females' average reflective ability was slightly higher than that of males (Kalelioğlu, 2015). Another study also showed that there was no significant difference between genders in LEGO construction and related programming, but females paid attention to the instructions of the task, whereas males rarely did (Lindh & Holgersson, 2007). Some studies have indicated significant gender differences in learning to program and acquiring computational thinking skills (Korkmaz & Altun, 2013; Özyurt & Özyurt, 2015).

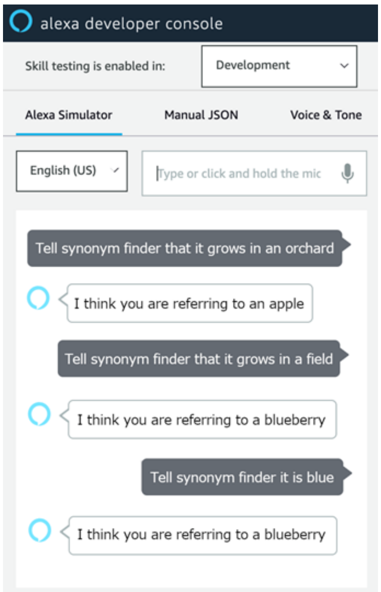
According to the cognitivist view of information processing theory, females tended to perceive information in detail and concentrate on sharing and correlating information when their brain processes the information, while males tended to pay attention to the information context (Putrevu, 2001). According to the selective input of information and the perspectives of gender schema in information processing theory, males and females have demonstrated slight differences in their methods of selecting and processing information.

Overview of the Study

Many countries have encouraged females to engage in STEM disciplines. Females' experiences during K–12 education affect their choices to continue with those subjects in the future. In addition, AI education in K–12 has become more popular (Long & Magerko, 2020; Touretzky et al., 2019). Due to this popularization and gender gap in STEM, it is important to explore the effects of gender on AI education. Specifically, we aimed to explore these effects using the conversational AI curriculum developed by Van Brummelen (2019). AI literacy has become increasingly important, particularly with the prevalence of voice-based AI technology such as Alexa, Google Home, Siri, and so on is. Voice technology is helpful for people who are not able to use conventional input devices, as they can directly talk to the computer or smartphone instead of typing or using a mouse. Figure 1 shows an example of a conversational AI application.

Figure 1

Example of a Conversational VoiceBot in the Alexa Simulator (Amazon, 2021)



The conversational AI system providing the voice user interface (VUI) is sometimes also called a voicebot, and is an intelligent assistant for humans' daily life, which interacts with people through voice conversations. Conversational AI is the skeuomorphism of VUI. The innovation in this study was to implement the conversational AI curriculum in the formal classroom setting of a secondary school. The two approaches used to instruct this conversational AI curriculum involved the cycle of doing projects and the cycle of experiential learning. It was expected that the junior high school students would gain hands-on experience of programming and the application of AI in the conversational AI curriculum.

The curriculum taught students to develop mobile applications and Amazon Alexa skills, the programs that run on voice-first Alexa devices, using MIT App Inventor (Van Brummelen, 2019). MIT App Inventor, a block-based programming tool that encouraged the practice of computational thinking, included logical and problem-solving processes. Our study evaluated whether different learning approaches (conventional instruction using the cycle of doing projects vs. the cycle of experiential learning) and different genders would have effects on the learning effectiveness of conversational AI, the performance of VUI, and the computational thinking concept scale of the students. The following research questions guided our investigation.

1. Does gender (i.e., males and females) and learning approach (i.e., cycle of doing projects and cycle of experiential learning) affect the learning effectiveness of the conversational AI curriculum?

2. Does gender and learning approach affect VUI performance in the conversational AI curriculum?

3. Does gender and learning approach affect students' understanding of computational thinking concepts?

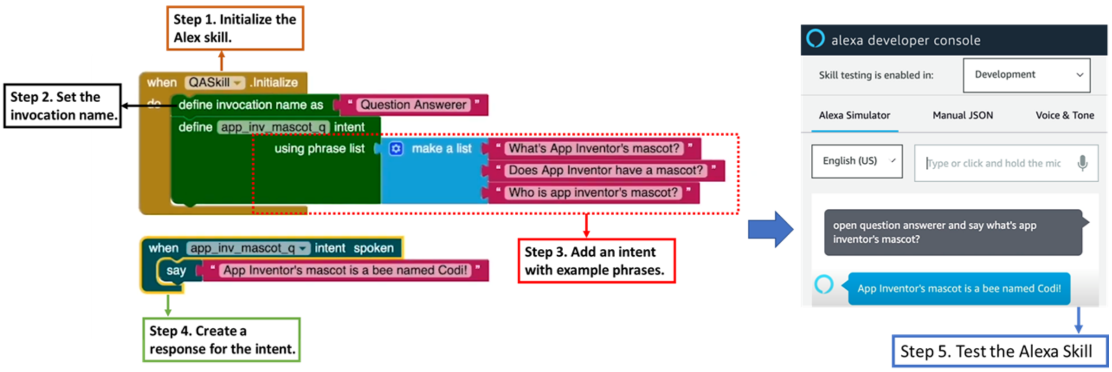
Method

Learning Conversational AI

The conversational AI used in this study involved using audio to control Amazon Alexa. To make an Alexa skill, the student learned to write the conversation program with block-based programming. First, the student logged onto MIT App Inventor, and initialized the Alexa skill by dragging from the block menu, shown as step 1 in Figure 2. Second, the student dragged-and-dropped blocks to program the Alexa skill, shown as steps 2 to 4 in Figure 2. Then the student clicked a button to send the skill to Amazon. Finally, this action converted the blocks into text-based code, which was readable by Alexa devices or the Amazon Website as shown in the right-hand screenshot in Figure 2.

Figure 2

Example Program in the Block-Based Programming Conversational AI Interface

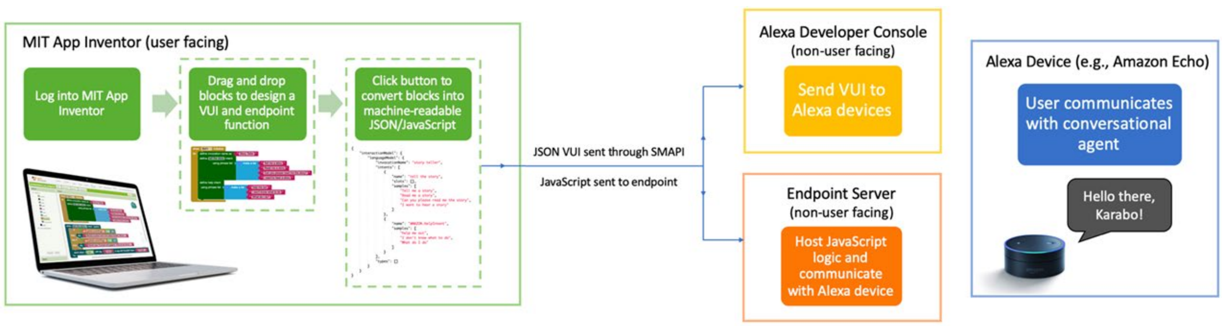


This conversational AI tool in the block-based programming environment was developed for K–12 students to create their own conversational agents (Van Brummelen, 2019). Students chatted with Alexa or the Alexa simulator Website after they wrote the conversational AI program. Amazon has embedded natural language processing inside their Alexa system and simulator. The combination of Alexa in Amazon and MIT App Inventor was chosen as a friendly learning tool and resource for primary or secondary school students to experience and apply conversational AI, even though they were not undergraduates in computer sciences.

The system framework behind the block-based programming platform is depicted in Figure 3. The system ensured low barriers to entry for primary and secondary school students, otherwise, creating Alexa skills would be difficult, even for a student majoring in computer science. For example, without the interface, connecting a lambda function on AWS to the voice user interface is complicated. However, the block-based interface design in Figure 3 abstracted that, and simplified the development of students' own conversational agent.

Figure 3

System Framework of the Conversational AI Programming Tool in MIT App Inventor



From "Tools to Create and Democratize Conversational Artificial Intelligence," by J. Van Brummelen, 2019, master's thesis, MIT, Cambridge, p. 52 (https://hdl.handle.net/1721.1/122704).

Conversational AI is directly related to Brennan and Resnik's (2013) computational thinking (CT) skill framework. In our study, students engaged with: (a) CT concepts including events, conditionals, data, sequences, loops, parallelism, and operators; (b) CT practices such as being incremental and iterative, testing and debugging, reusing and remixing, as well as abstracting and modularizing; and (c) CT perspectives like expressing, connecting, questioning, and so on. In addition to computational thinking naturally embedded in the conversational AI curriculum, students also learned AI-specific concepts, practices, and perspectives including, but not limited to (a) classification (e.g., determine intent); (b) prediction (e.g., predict best next letter); (c) generation (e.g., generate text block); (d) training, testing, and validating (e.g., vary training length); and (e) project evaluation (e.g., question project ethics).

Two Approaches to Learning Conversational AI

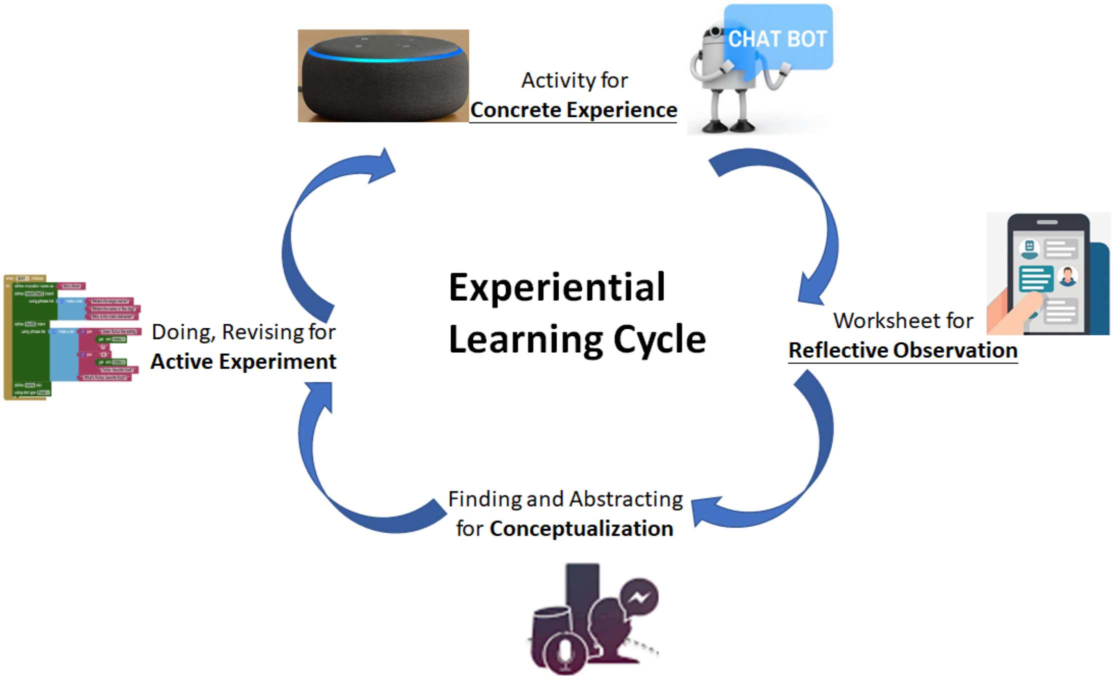
It was hypothesized that an appropriate instructional approach will be helpful for assisting the students in learning to make conversational AI in computer education. Therefore, this empirical study aimed to evaluate two different learning approaches in two classes, respectively.

The Experimental Group: Experiential Learning

One class, labelled the experimental group, used the cycle of experiential learning; its instructional design is exhibited in Figure 4. The students already had concrete experience using conversational AI. For example, they used the phrase “Hey Google” to give their mobile phone oral rather than text commands, so that they could receive the oral and data response of the smartphone. The students filled out a worksheet about what they observed and found after they used conversational AI in their daily life. At this stage, they also thought about new tasks. The teacher encouraged the students to have conversations with the computer, and the students filled out the worksheet to show what they said and how the system reacted. The students also practiced problem decomposition in this stage. After the students progressed to the abstract conceptualization stage, they practiced pattern recognition and abstraction for problem solving. At this stage, students used their Amazon account to log into MIT App Inventor, but they did not yet write their own program. The teacher provided them with different blocks, and asked them to conceptualize which block could be used for which task. Finally, in the active experiment stage, the students actually implemented their own program and tested the running results. If they encountered any problems, they debugged and revised the program. During the process, they asked the teacher questions if they had a problem.

Figure 4

Experiential Learning Cycle Integrated into the Experimental Group's Learning Process

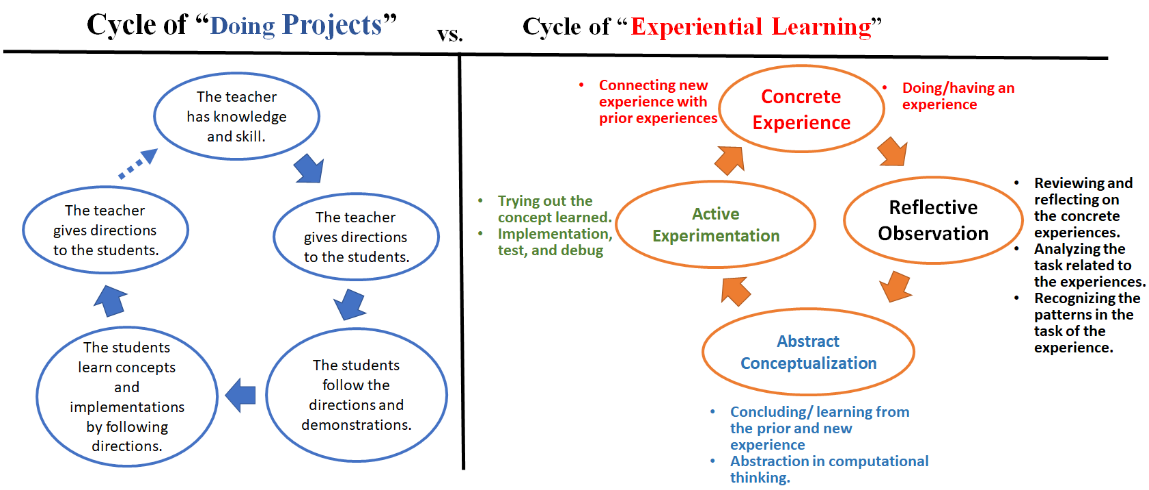


Control Group: Cycle of Doing Projects

The conventional instruction approach, referred to as the cycle of doing projects, was used in the other class and is depicted on left-hand side of Figure 5. The teacher guided the process step-by-step. Students followed the teacher's directions and when they implemented the project of conversational AI, students imitated the teacher's demonstration of the codes. The difference between the cycle of doing projects in the conventional instruction of this study and the cycle of experiential learning is compared and illustrated in Figure 5.

Figure 5

The Cycle of Doing Projects (Control Group) Compared With the Cycle of Experiential Learning (Experimental Group)



Participants

A total of 46 seventh-grade students participated in the conversational AI curriculum. As shown in Table 1, 25 were assigned to the experimental group and experiential learning, and 21 were assigned to the control group and the general cycle of doing projects.

Table 1

Gender and Number of Participants in Learning Approach Groups

Learning approach	Gender	*n*
Cycle of experiential learning (experimental group)	Female	11
	Male	14
Cycle of doing projects (control group)	Female	7
	Male	14

The purpose of this research study was to determine whether the students could understand conversational artificial intelligence (the ability for a computer to have conversations with humans) and develop programming projects through formal classes in secondary school via two different learning approaches. Participation in the study was completely voluntary and the students' parents filled out the consent form. The students were able to decline to answer any or all of the questions. If a student declined to answer any of the questions, he or she would no longer be participating in the study. The students could decline participation at any time. The data collected in this study were reported in a that protected individuals' identities.

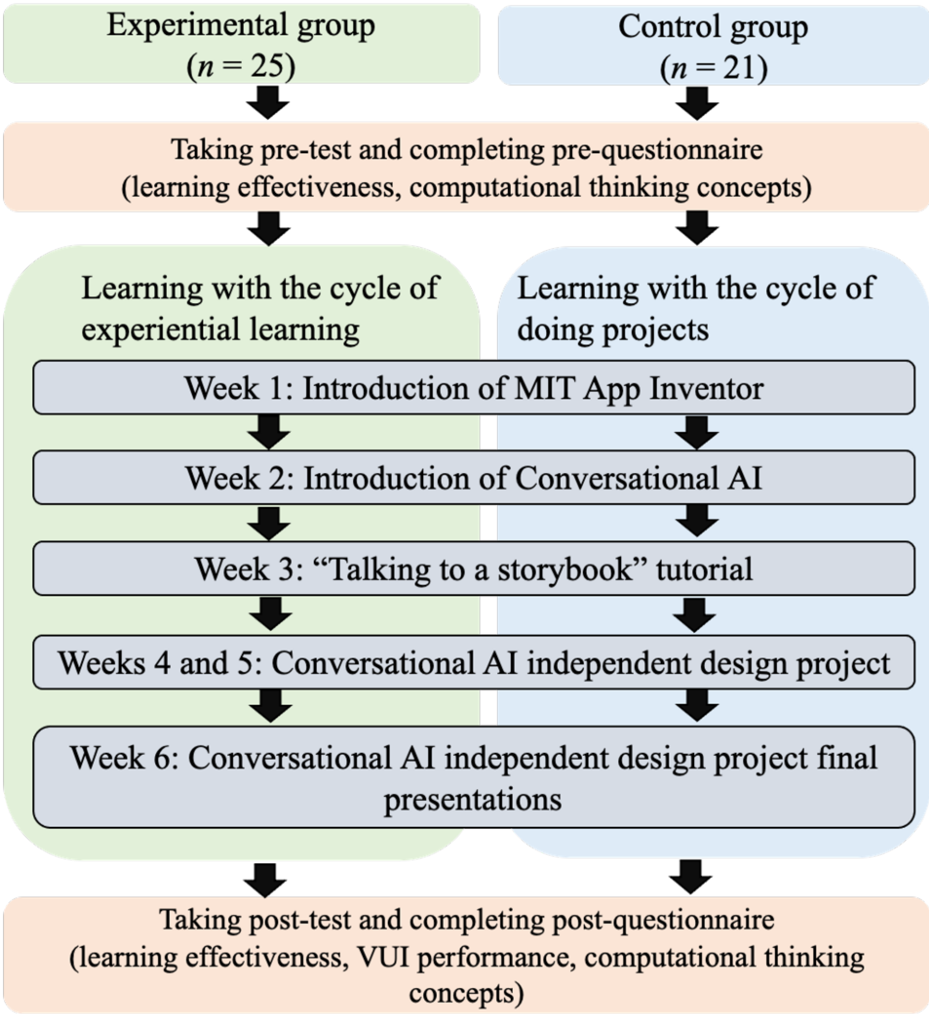
Experimental Process and Measuring Tools

The conversational AI curriculum took a total of six weeks. The students in the two classes learned computational thinking and AI skills from the curriculum after they developed their own conversational AI projects during the six weeks. The learning objectives of the conversational AI curriculum were to learn how conversational agents decide what to say, to comfortably develop the conversational AI projects, and to better understand conversational agents. Accordingly, the students were encouraged to develop positive, socially useful, and meaningful projects in the course.

The pre-test of prior knowledge included 15 multiple-choice questions, with a perfect score of 100. The post-test for measuring the learning effectiveness also comprised 15 multiple-choice questions, with a perfect score of 100.

Figure 6

The Experimental Flow Chart



The VUI performance and computational thinking concepts were measured with a five-point Likert scale, ranging from *strongly disagree* to *strongly agree*. The VUI performance scale had five questions (Van Brummelen, 2019), namely, (a) I have interacted with conversational agents, (b) I understand how conversational agents decide what to say, (c) I feel comfortable making apps that interact with conversational agents, (d) I can think of ways that conversational agents can solve problems in my everyday life, and (e) my understanding of conversational agents improved through the curriculum. The Cronbach's α value of the reliability of the VUI performance scale was 0.883. The computational thinking concept scale had five questions (Sáez-López et al., 2016), outlined below.

After learning block-based programming, I:

1. understand sequences with combined characters, backgrounds, and elements
2. can include loops in programming to allow a proper multimedia product
3. can add parallelism and events that allow the creation of interfaces
4. have an improved ability to share and play with the content created
5. acquired the ability to communicate and express through the content created

The reliability of the original combined scale was 0.789. The Cronbach's α value of the retest reliability of the computational thinking concept scale was 0.921.

The students' behaviors were video-recorded in the class. After the quantitative analysis, the recordings were used to infer and understand why the students learned well or not.

Results

Learning Effectiveness of Different Learning Approaches With Different Genders

Two-way ANCOVA was employed to compare the learning effectiveness of the conversational AI curriculum with different learning approach (i.e., the cycle of doing projects and the cycle of experiential learning) and gender (males and females). The covariance was the pre-test used to measure the prior knowledge of the students before the conversational AI curriculum. The independent variables were gender and the learning approach. The dependent variable the post-test used to measure the students' learning effectiveness after they completed the curriculum. The Levene's test was not violated ($F = 1.424$, $P = .249 > .050$), suggesting that a common regression coefficient was appropriate for the two-way ANCOVA.

Table 2 shows the two-way ANCOVA results. It was found that there was interaction between the two independent factors, learning approach and gender, for the students' learning results ($F = 12.493^{**}$, $P = .001 < .010$). The effect size (partial η^2) was 0.247, indicating a medium effect.

Table 2

Two-Way ANCOVA Tests of Between-Subjects Effects

Resource	*SS*	*MS*	*F*	*P*	Partial η^2
Learning approach * Pre-test	362.82	362.82	0.929	.341	
Gender * Pre-test	898.18	898.18	2.300	.138	
Learning approach	117.24	117.24	0.300	.587	
Gender	83.65	83.65	0.214	.646	
Learning approach * Gender	4879.23	4879.23	12.493**	.001	0.247

Note. ** $p < .01$.

A simple main-effect analysis based on the division of gender was explored; results are presented in Table 3. When the group was divided based on gender, the Levene's test was not violated for males (F = 0.086, P = .772 > .050) or females (F = 2.137, P = .163 > .050). However, the pre-test had interaction with learning approach for males ($F = 4.803^{*}$; P = .038 < .050) as well as females ($F = 8.012^{*}$; P = .013 < .050). Therefore, the Johnson-Neyman process was further conducted.

Table 3

Simple Main-Effect Analysis Based on the Division of Gender

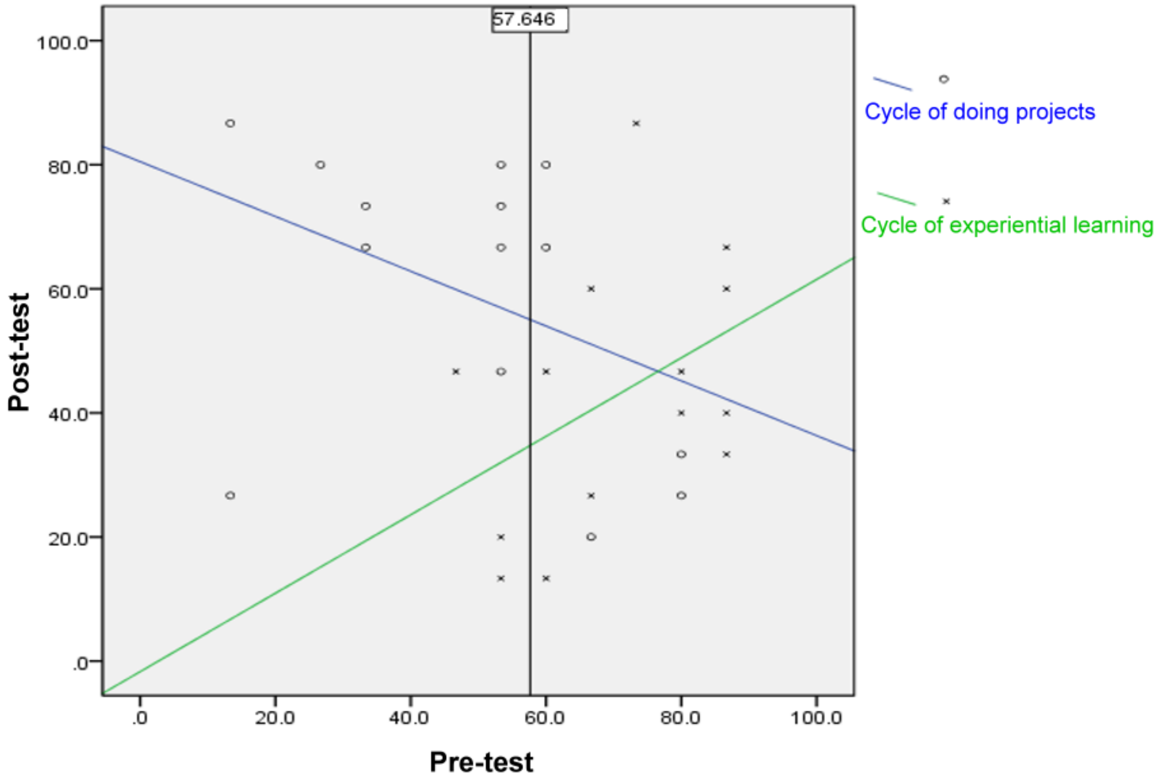
Gender	Learning Approach	*n*	Mean	*SD*	Adjusted mean	*SE*
Female	Cycle of experiential learning	11	71.52	19.34	71.48	6.76
	Cycle of doing projects	7	67.84	8.50	67.84	8.50
Male	Cycle of experiential learning	14	42.86	21.04	42.15	6.47
	Cycle of doing projects	14	59.05	23.37	59.68	6.37

Note. * p < .05.

For males, it was found that when the pre-test was smaller than 57.646, the male students using the cycle of doing projects outperformed the male students using the cycle of experiential learning, as shown as Figure 7. Conversely, the high-prior competence of the males using the cycle of experiential learning performed better than the high-prior competence of the males using the cycle of doing projects.

Figure 7

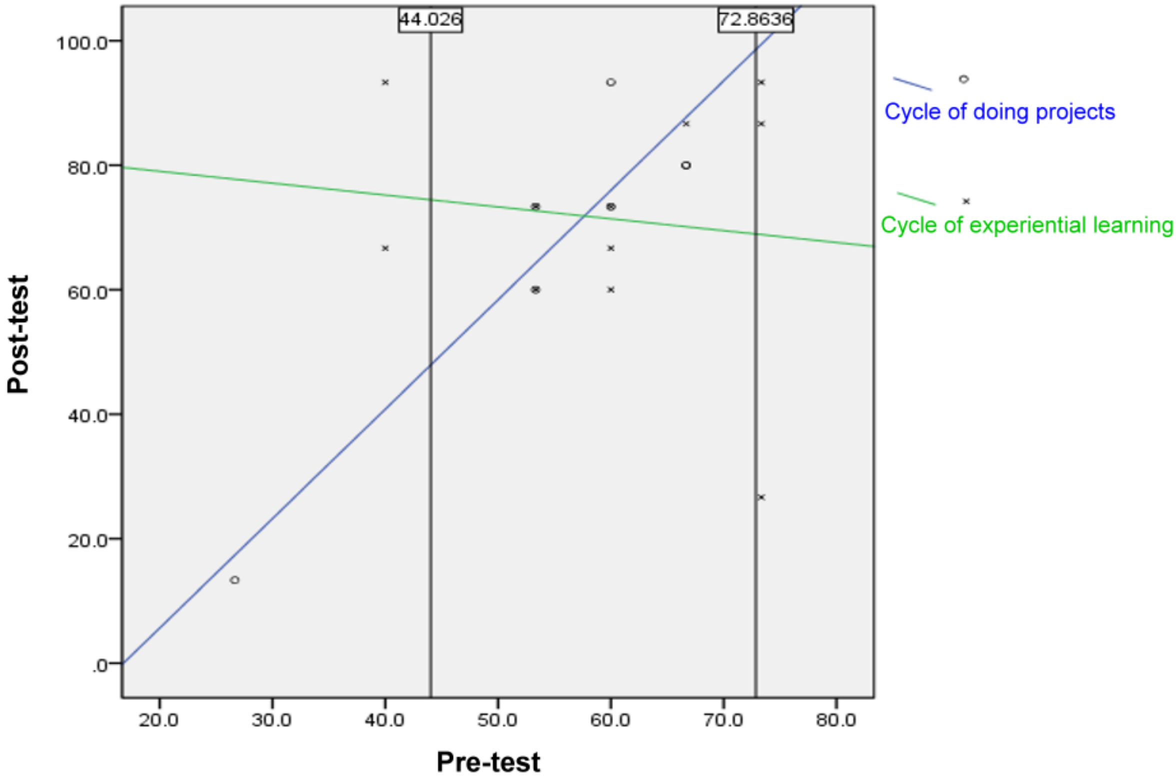
Results of Johnson-Neyman Process for Males Using Different Learning Approaches



For females, when the pre-test was smaller than 44.026, the female students using the cycle of experiential learning outperformed the female students using the cycle of doing projects. Conversely, when the pre-test was larger than 72.864, the female students using the cycle of doing projects performed better than the female students using the cycle of experiential learning, as shown as Figure 8.

Figure 8

Results of Johnson-Neyman Process for Females Using Different Learning Approaches



A simple main-effect analysis based on the division of learning approaches was further explored; see results in Table 4. When the group was divided based on learning approach, the Levene's test was not violated for the cycle of experiential learning approach (F = 0.116, P = .737 > .050) or for the cycle of doing projects (F = 4.101, P = .057 > .050). The pre-test had no interaction with gender for the cycle of experiential learning approach (F = 1.596; P = .220>.050). However, the pre-test had interaction with gender for the cycle of doing projects (F = 12.146**; P = .003 < .010). Therefore, the Johnson-Neyman process was further conducted.

Table 4

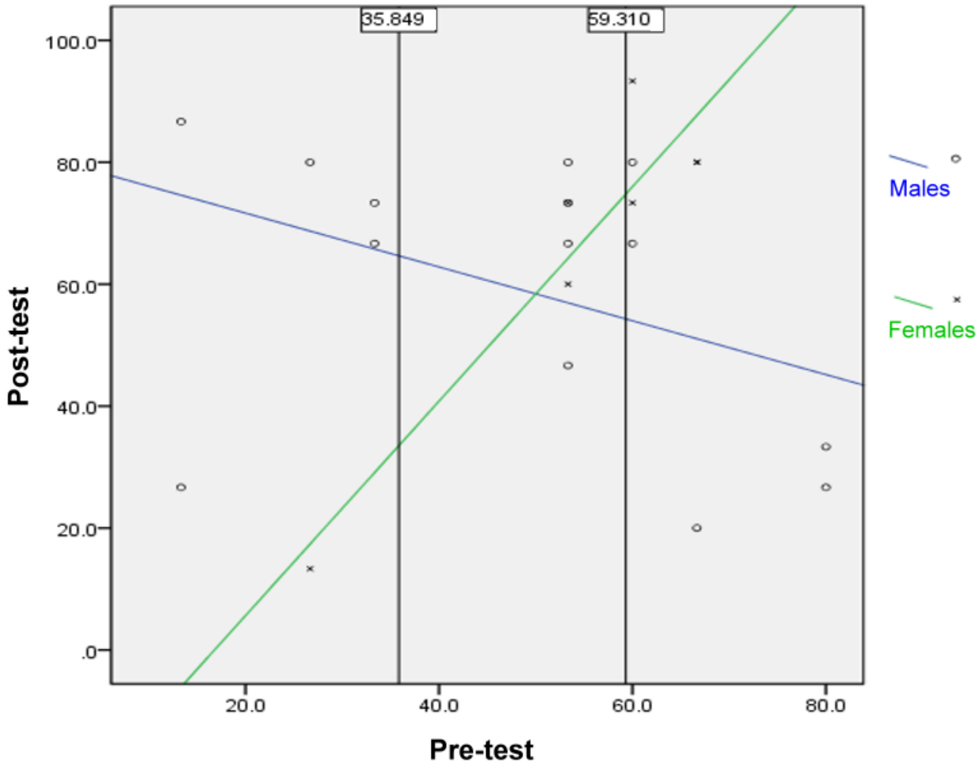
Simple Main-Effect Analysis Based on the Division of Learning Approaches

Learning approach	Gender	*n*	Mean	*SD*	Adjusted mean	*SE*
Cycle of experiential learning	Female	11	71.52	19.34	71.48	6.76
	Male	14	42.86	21.04	42.15	6.47
Cycle of doing projects	Female	7	67.62	25.94	67.84	8.50
	Male	14	59.05	23.37	59.68	6.37

As for the group using the cycle of doing projects, when the pre-test was smaller than 35.849, the males outperformed the females. Conversely, when the pre-test was larger than 59.310, the females performed better than the males, as shown as Figure 9.

Figure 9

Results of Johnson-Neyman Process for Males and Females Using the Cycle of Doing Projects



Consequently, instructors are advised consider students' prior knowledge when they choose learning approaches for the secondary school students learning conversational AI. Overall, the cycle of experiential learning was as effective as the cycle of doing projects for this curriculum. However, there was a significant interaction between gender and learning approach. From the classroom observations, this study found that most of the males tended to be distracted when they first studied the AI curriculum.

VUI Performance of Different Genders With Different Learning Approaches

There were five items in the questionnaire of the performance of VUI. A two-way ANOVA was used to analyze the average scores of the five items determining whether students understood conversational artificial intelligence and had developed programming projects through the formal class in the secondary school with two different learning approaches. The dependent variable was the survey results after the instructional experiment. The two independent variables were gender and learning approach. The Levene's test of determining homogeneity of regression was not violated ($F(3,42) = 1.303$, $P = .286 > .05$).

Table 5 shows the two-way ANOVA results of the VUI performance. It was found that there was significant impact on the interaction between learning approach and gender ($F = 4.581^{*}$, $P = .035 < 0.05$, partial $\eta 2 = 0.098$). At the same time, it was found that there were significant effects for gender ($F =$

6.543*, P = .014 < 0.05, partial η2 = 0.135) on students' perspectives of the conversational AI curriculum, while no significant effect was found for students' perspectives in the different learning approach conditions (F = 0.330, P = .569 > .05).

Table 5

Tests of Between-Subject Effects Measure in the Two-Way ANOVA for VUI Performance

Source factor	Type III *SS*	*MS*	*F*	*P*	Partial η^2
Learning approach	0.262	0.262	0.330	.569	
Gender	5.204	5.204	6.543*	.014	0.135
Learning approach * Gender	3.644	3.644	4.581*	.038	0.098

Note. * p < .05.

Because there was interaction between students' VUI performance in the different learning approach conditions and for the different genders, simple main-effect analysis was further conducted. From the results presented in Table 6, we see that the VUI performance of the females learning with the cycle of experiential learning (mean = 4.00; *SD* = 0.63) and the cycle of doing projects (mean = 4.43; *SD* = 0.63) was similar (t = 1.416; P = .176 > .05). Furthermore, no significant difference (t = 1.924, P = .065 > .050) was found between the perspectives of males with the cycle of experiential learning (mean = 3.88; *SD* = 1.03) and the cycle of doing projects (mean = 3.14; *SD* = 1.01). In the cycle of doing projects, females' VUI performance (mean = 4.43; *SD* = 0.63) outperformed males' (mean = 3.14; *SD* = 1.01), which resulted in a significant difference (t = 2.923**; P = .009 < .01) with an effect size of 1.53. For the experiential learning approach, no significant difference (t = 0.322; P = .750 > .050) was found between the VUI performance of females (mean = 4.00; *SD* = 0.63) and males (mean = 3.88; *SD* = 1.03). Overall, the VUI performance of the females outperformed that of the males.

Table 6

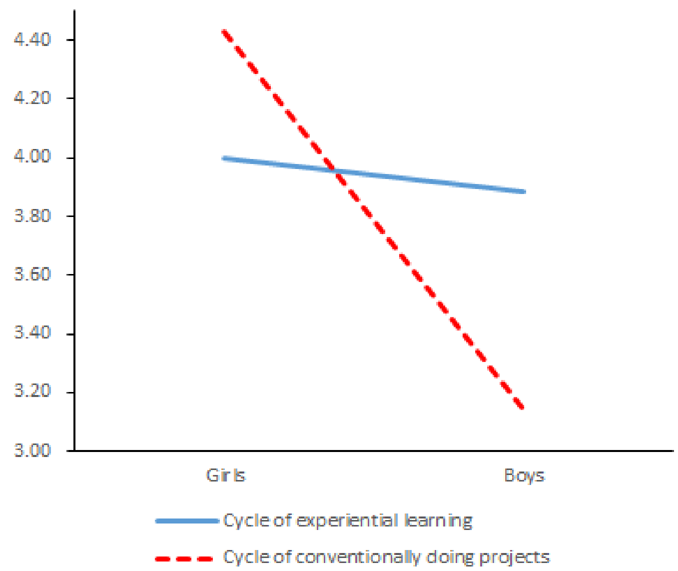
Descriptive Statistics Results After the Simple Main-Effect Analysis in VUI Performance

Learning approach	Gender	*n*	Mean	*SD*	Adjusted mean	*SE*
Cycle of experiential learning	Female	11	4.00	0.63	4.00	0.27
	Male	14	3.88	1.03	3.89	0.24
Cycle of doing projects	Female	7	4.43	0.63	4.43	0.34
	Male	14	3.14	1.01	3.14	0.24

Figure 10 shows the interaction between learning approach and gender on the students' VUI performance. In the cycle of doing projects, the VUI performance of females was significantly better than that of males.

Figure 10

Interaction Between Learning Approach and Gender Regarding Students' VUI Performance



Students' Computational Thinking with Different Learning Approaches and Gender

The two-way ANCOVA was employed to compare the computational thinking of students using the different instructional approaches and their gender. The covariance was the initial measurement of computational thinking before the learning activity took place. The independent variables were gender (i.e., male and female) and learning approach (i.e., experiential learning and project-based learning). The dependent variable was the post-measurement of the computational thinking scale. Levene's test was not violated ($F(3,42) = 0.636$, $P = .596 > .050$), suggesting that a common regression coefficient was appropriate for the two-way ANCOVA.

Table 7 shows the two-way ANCOVA results on the computational thinking scale. It was found that the covariance (i.e., the pre-measurement of computational thinking) would not cause significant effects on the interaction between the two factors, namely learning approach and gender, for the students' computational thinking concepts. Therefore, it was meaningful to directly examine the interaction between learning approach and gender on students' computational thinking. When the pre-measurement was not taken into consideration in the interaction, there was significant interaction between the two independent variables ($F(3,42) = 7.047^{*}$, $p = .011 < 0.050$). Furthermore, the effect size (partial η^2) of the interaction between learning approach and gender was 0.147, indicating a small to medium effect, larger than 0.10 presenting a small effect (Cohen, 1988).

Table 7

Two-Way ANCOVA Tests of Between-Subjects Effects on Computational Thinking Concepts

Resource	*SS*	*MS*	*F*	*P*	Partial η²
Learning approach * Pre-test	0.12	0.12	0.200	.658	
Gender * Pre-test	2.02	2.02	3.438	.071	
Learning approach * Gender * Pre-test	0.53	0.53	0.896	.350	
Learning approach	0.00	0.00	0.000	.992	
Gender	0.33	0.33	0.537	.468	
Learning approach * Gender	4.30	4.30	7.047^{*}	.011	0.147

Note. * p < .05.

Because the interaction between learning approach and gender was significant, simple main-effect analysis was used. Table 8 shows that the computational thinking of males with the experiential learning approach (mean = 3.86; *SD* = 0.91) outperformed ($t = 2.140^{*}$; $P = .042 < 0.50$) that of the males with the cycle of doing projects (mean = 3.19; *SD* = 0.74), with an effect size of 0.81. With the conventional instruction of the cycle of doing projects, females (mean = 4.20; *SD* = 0.77) presented significantly ($t = 3.066^{**}$; $P = .006 < .010$) better computational thinking than did males (mean = 3.19; *SD* = 0.74) with an effect size of 1.34. There was no significant difference ($t = 1.791$, $P = .095 > .050$) between the computational thinking of females with the cycle of experiential learning approach (mean = 3.51; *SD* = 0.85) or the cycle of doing projects (mean = 4.20; *SD* = 0.77).

Table 8

Descriptive Data after the Simple Main-Effect Analysis for Computational Thinking Concepts

Learning approach	Gender	*n*	Mean	*SD*	Adjusted mean	*SE*
Cycle of experiential learning	Female	11	3.51	0.85	3.44	0.24
	Male	14	3.86	0.91	3.90	0.21
Cycle of doing projects	Female	7	4.20	0.77	4.08	0.30
	Male	14	3.19	0.74	3.26	0.21

Discussion and Conclusion

According to the results of this empirical study, when teachers instruct secondary school students to learn conversational AI curriculum, it is recommended that the low-achievement males and high-achievement females adopt the cycle of doing projects. It is also suggested that the high-achievement males and low-achievement females use the cycle of experiential learning, so as to meet their individual needs and differences.

This empirical study of applying the conventional cycle of doing projects to conversational AI curriculum found that females performed better than males in terms of computational thinking concepts. Based on information processing theory in cognitivism, males and females do not have the same level of focus when receiving and processing information. According to this theory, males require strong context linkage when processing information; we suggest that instructors provide additional scaffolding. It particular, it would be helpful to focus on context for male, so as to prevent them from being distracted, as was found in this study.

According to information processing theory, females focus on sharing information and developing correlations among the information they are aware of. In comparison with males, females are accustomed to taking in detailed information and understanding detailed processes. Therefore, in the future, it will be important to explore further the effects of various learning approaches on K–12 students of different gender learning AI.

Limitations of this study included the sample size for the instructional experiments, and the small number of the countries with experience learning the new functions of conversational AI in MIT App Inventor. Due to increased use of reliance on IoT, future research to apply the conversational AI tool used in this current study to K–12 education is encouraged.

Statements on Open Data, Ethics, and Conflict of Interest

The dataset is available by contacting the corresponding author.

The ethics rules and regulations of the Declaration of Helsinki were followed during the experiment. All the participants were volunteers and were told that they could quit the study at any time.

The researchers have no conflict of interest.

Acknowledgements

This study was supported in part by the Ministry of Science and Technology under contract numbers MOST 108-2511-H-003 -056 -MY3 and by the Hong Kong Jockey Club Charities Trust.

References

Adya, M., & Kaiser, K. M. (2005). Early determinants of women in the IT workforce: A model of girls' career choices. *Information Technology and People, 18*(3), 230–259. https://doi.org/10.1108/09593840510615860

Amazon. (2021). Alexa Developer Console. Retrieved October 20, 2021, from https://developer.amazon.com/alexa/console/ask

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 Annual Meeting of the American Educational Research Association* (pp. 1–25). Vancouver, Canada.

Cetin, I. (2016). Preservice teachers' introduction to computing: Exploring utilization of Scratch. *Journal of Educational Computing Research*, *54*(7), 997–1021. https://doi.org/10.1177/0735633116642774

Cheng, G. (2019). Exploring factors influencing the acceptance of visual programming environment among boys and girls in primary schools. *Computers in Human Behavior, 92*, 361–372. https://doi.org/10.1016/j.chb.2018.11.043

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*(1), 1–35.

Chiu, C. F. (2020). Facilitating K–12 teachers in creating apps by visual programming and project-based learning. *International Journal of Emerging Technologies in Learning*, *15*(1), 103–118.

Efstratia, D. (2014). Experiential education through project based learning. *Procedia - Social and Behavioral Sciences, 152*, 1256–1260. https://doi.org/10.1016/j.sbspro.2014.09.362

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher, 42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Hsu, T.-C., Abelson, H., Lao, N., Tseng, Y.-H. & Lin, Y.-T. (2021). Behavioral-pattern exploration and development of an instructional tool for young children to learn AI. *Computers and Education: Artificial Intelligence, 2*, 100012. https://doi.org/10.1016/j.caeai.2021.100012

Hsu, T.-C., Chang, S.-C., & Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education, 126*, 296-310. https://doi.org/10.1016/j.compedu.2018.07.004

Kalelioğlu, F. (2015). A new way of teaching programming skills to K–12 students: Code. org. *Computers in Human Behavior, 52*, 200–210. https://doi.org/10.1016/j.chb.2015.05.047

Kelleher, C., & Pausch, R. (2005). Lowering the barriers to programming: A taxonomy of programming environments and languages for novice programmers. *ACM Computing Surveys, 37*(2), 83–137. https://doi.org/10.1145/1089733.1089734

Korkmaz, Ö., & Altun, H. (2013). Engineering and CEIT student's attitude towards learning computer programming. *International Journal of Social Science Studies, 6*(2), 1169–1185.

Lindh, J., & Holgersson, T. (2007). Does LEGO training stimulate pupils' ability to solve logical problems? *Computers & Education, 49*(4), 1097–1111. https://doi.org/10.1016/j.compedu.2005.12.008

Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery, New York, NY, USA.

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K–12? *Computers in Human Behavior, 41*, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin, 128*(6), 903–933.

Meyers-Levy, J. (1986). *Gender differences in information processing: A selectivity interpretation*. Northwestern University.

Meyers-Levy, J. (1989). The influence of a brand name's association set size and word frequency on brand memory. *Journal of Consumer Research*, *16*(2), 197–207. https://doi.org/10.1086/209208

Özyurt, Ö., & Özyurt, H. (2015). A study for determining computer programming students' attitudes towards programming and their programming self-efficacy. *Journal of Theory and Practice in Education, 11*(1), 51–67.

Piwek, P. & Savage, S. (2020). Challenges with learning to program and problem solve: An analysis of student online discussions. *The 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20* (pp. 494–499). Association for Computing Machinery, Portland, OR, USA. https://doi.org/10.1145/3328778.3366838

Pucher, R. & Lehner M. (2011). Project based learning in computer science – A review of more than 500 projects. *Procedia - Social and Behavioral Sciences, 29*, 1561–1566. https://doi.org/10.1016/j.sbspro.2011.11.398

Putrevu, S. (2001). Exploring the origins and information processing differences between men and women: Implications for advertisers. *Academy of Marketing Science Review*, *10*(1), 1–14.

Sáez-López, J. M., Román-González, M., & Vázquez-Cano, E. (2016). Visual programming languages integrated across the curriculum in elementary school: A two year case study using "Scratch" in five schools. *Computers & Education, 97*, 129-141. https://doi.org/10.1016/j.compedu.2016.03.003

Sendall, P., Stuetzle, C. S., Kissel, Z. A., Hameed, T. (2019). Experiential learning in the technology disciplines. *Proceedings of the 2019 EDSIG Conference* (Vol.5, n4968). Information Systems and Computing Academic Professionals.

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019, July). Envisioning AI for K-12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 9795–9799). https://doi.org/10.1609/aaai.v33i01.33019795

Van Brummelen, J. (2019). *Tools to create and democratize conversational artificial intelligence* [Master's thesis, Massachusetts Institute of Technology]. https://hdl.handle.net/1721.1/122704

Weston, T. J., Dubow, W. M., Kaminsky, A. (2019). Predicting women's persistence in computer science- and technology-related majors from high school to college. *ACM Transactions on Computing Education, 20*(1). https://doi.org/10.1145/3343195

Yukselturk, E., & Altiok, S. (2017). An investigation of the effects of programming with Scratch on the preservice IT teachers' self-efficacy perceptions and attitudes towards computer programming. *British Journal of Educational Technology, 48*(3), 789–801. https://doi.org/10.1111/bjet.12453

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 23, Number 1

February – 2022

AI in Online-Learning Research: Visualizing and Interpreting the Journal Publications from 1997 to 2019

Gwo-Jen Hwang[1], Yun-Fang Tu[2], and Kai-Yu Tang[3]
[1]National Taiwan University of Science and Technology, [2]Fu Jen Catholic University, [3]Ming Chuan University

Abstract

This study reviews the journal publications of artificial intelligence-supported online learning (AIoL) in the Web of Science (WOS) database from 1997 to 2019 taking into account the contributing countries/areas, leading journals, highly cited papers, authors, research areas, research topics, roles of AIoL, and adopted artificial intelligence (AI) algorithms. Results indicate that, from 1997 to 2009, AIoL research focused on the combination of intelligent tutoring systems and distance learning. In 2010–2014, AIoL research emphasized learner-oriented learning. In 2015–2019, learner-system interactions to facilitate personalized, adaptive, and collaborative learning became the main focus. "Intelligent tutoring systems" have played the most important role in AIoL, followed by "profiling and prediction," and "adaptive systems with personalization." Accordingly, the roles and research trends as well as several suggestions for future research in the field of AIoL are provided as a reference for researchers and policy makers.

Keywords: artificial intelligence, online learning, literature review, trend analysis, visualization

Introduction

In past decades, online learning has been adopted by researchers and educators for delivering courses in various domains (Hwang & Tsai, 2011; Martin et al., 2020). Nevertheless, many researchers have pointed out the problems of this learning mode, including the low assignment completion rate and poor learning outcomes (Lee & Choi, 2011; Yu et al., 2017). Dropout rates are one of the thorny issues in online learning environments. Some scholars have attempted to reduce dropout rates through strategies such as understanding each student's challenges and potential, providing quality curricular activities with good learning supports, and promoting quality learning experiences with learning guidance (Lee & Choi, 2011; Hussain et al., 2019; Lee et al., 2020). However, it is nearly impossible for teachers to provide personalized learning support or guidance to individual learners when they need to face dozens, hundreds, or even thousands of students in online classes.

The advancement of artificial intelligence (AI) technologies provides an opportunity to address this problem. AI technologies can be used not only to predict students' learning status, but also to provide required support or guidance by analyzing students' online learning behaviors, personal characteristics (e.g., preferences or cognitive styles), and learning performances (Hwang et al., 2020). Several scholars have reported that using AI technologies to provide personalized learning supports has good potential to promote learner engagement (Lin et al., 2018) and enhance students' positive learning experiences (Yu et al., 2017). In addition, AI technology can be used to diagnose learners' personal learning problems and provide immediate assistance or advice accordingly (Chen et al., 2020; Chen & Lain, 2020).

With the growing interest in AI-supported education, researchers have conducted review studies on particular research domains, including medical education (Han et al., 2019), engineering education (Shukla et al., 2019), higher education (Zawacki-Richter et al., 2019), and e-learning research (Tang et al., 2021). In medical education, Han et al. (2019) proposed some thematic trends and explained the trends of advanced technology and artificial intelligence for future physicians. Using bibliometric analysis, Shukla et al. (2019) compared two reputed databases, the Web of Science and Scopus, and identified some frequently cross-referenced engineering applications of artificial intelligence. Zawacki-Richter et al. (2019) conducted a study to review AI in education research published between 2007 and 2018, and reported that AI technologies can facilitate profiling and prediction as well as the development of intelligent tutoring systems (ITS), adaptive systems, and recommendation systems. However, in their study, the features of online learning and several important characteristics of AI in education (e.g., contributing countries and authors, research areas, and research topics), which are valuable for providing clear direction to novice researchers, were not taken into account. Recently, researchers used co-citation network analysis to identify some highly co-cited research streams and their extensions in the e-learning area (Tang et al., 2021). They found that AI has been mainly used as an adaptive learning environment for learners. In addition to bibliometric analysis, some researchers have also proposed a state-of-the-art overview and positioning review to provide expert opinions to AI development, including machine learning for e-learning (Khanal et al., 2020) and deep learning in medical education (Carin, 2020). The main topics and methods of the previous review studies are summarized in Table 1.

Table 1

A List of Recent AI-Related Review Studies With an Education Focus

Year	Reviewed topics	Review method
2019	Advanced technology and AI in medical education	Bibliometric analysis (2010–2019)
2019	Engineering applications of AI	Bibliometric analysis (1988–2018)
2019	Research on AI applications in higher education	Bibliometric analysis (1988–2018)
2020	AI-supported e-learning	Systematic review and co-citation network analysis (1998–2019)
2020	AI and deep learning in medical education	Positioning review of deep learning
2020	Machine learning for e-learning	Trending overview of the research states and remaining challenges

The above-mentioned research suggests that a review study would be valuable and could help novice researchers efficiently and effectively acquire knowledge of the research trends and focus in the field. It also implied that it would be important to conduct review studies for AI in online learning (AIoL) since AIoL is becoming an important field of educational technology (Chen & Lain, 2020). Taken together, this present study aimed to contribute to the AIoL literature in two ways. First, a series of bibliometric analyses provided quantitative results to represent the international publication patterns of AIoL research, including the most productive countries, journals, highly cited papers, and authors in the field. The results could complement some conceptual frameworks of previous narrative reviews and positional papers. In addition to the bibliometric results, this study visualized the main characteristics of AIoL research, such as the most frequently referenced keywords and research topics. The visualization results not only present a holistic picture of the field but also provide a structural understanding of the most influential relationships of AIoL research patterns.

To gain an in-depth understanding of the roles and trends of AIoL research, this study aimed to review relevant journal publications from the WoS (Web of Science) database. In this context, with the aims of providing a guide for new research, identifying trends in the field, and comparing existing research on the topic, the following research questions were addressed:

- What were the major countries/areas conducting AIoL research in 1997–2019?
- What/who were the leading journals, papers, and authors of AIoL in 1997–2019?
- What was the distribution of the main research areas of AIoL applications in 1997–2019?
- What were the research topics of AIoL research in 1997–2019?
- What were the roles of AIoL and adopted AI algorithms in AIoL in 1997–2019?

Method

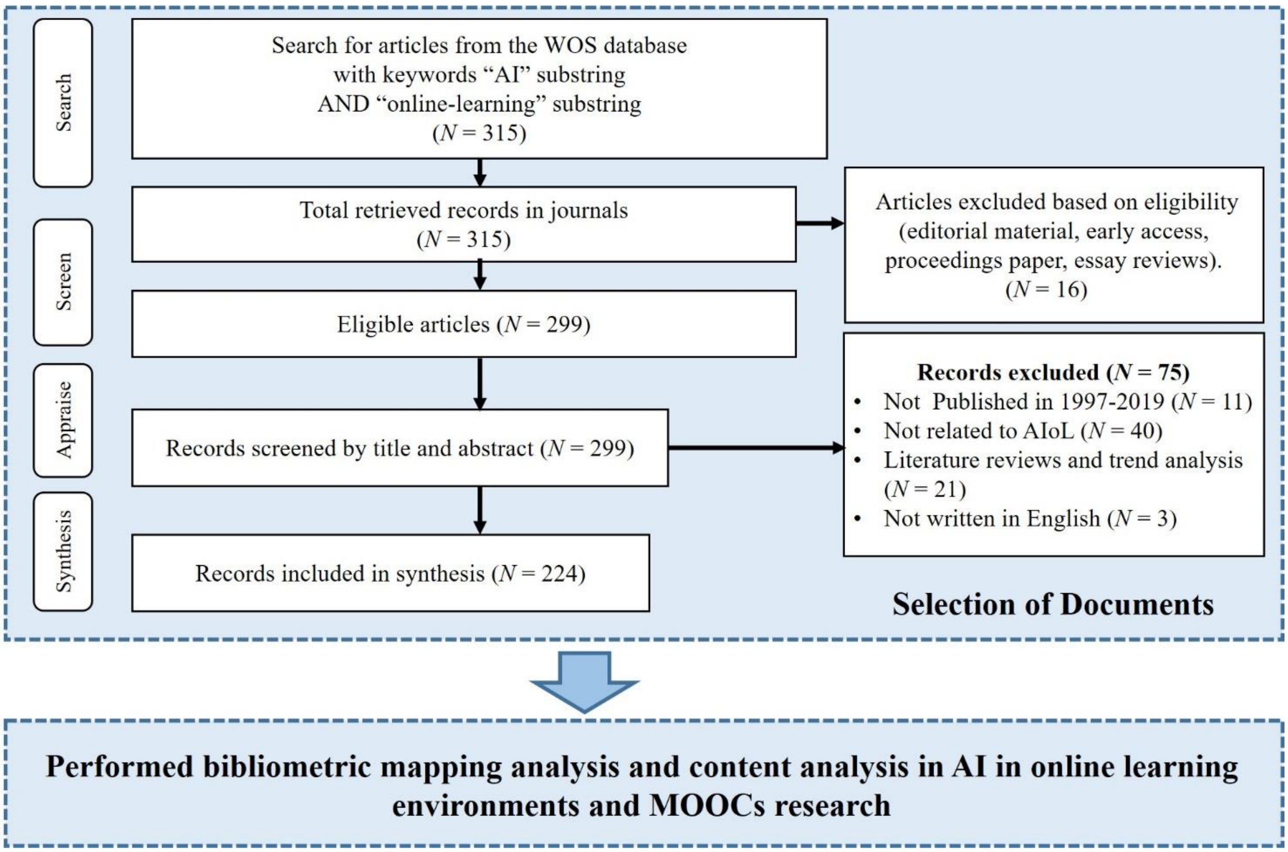
Article Selection Process and Data Coding

The current study was based on the research purpose and referred to the literature reviews on AIoL by Shukla et al. (2019), Wong et al. (2019), and Zawacki-Richter et al. (2019). First, on October 31, 2020, we searched for papers published in the WOS database. Based on the searching list of Social Sciences Citation Index (SSCI) journals, until the end of 2019, there were 315 articles including AI ("artificial intelligence" or "machine intelligence" or "intelligent support" or "intelligent virtual reality" or "chat bot*" or "machine learning" or "automated tutor*" or "personal tutor*" or "intelligent agent*" or "expert system*" or "neural network*" or "natural language processing" or "chatbot*" or "intelligent system" or "intelligent tutor*") AND online learning ("online learning" or "e-learning" or "Internet learning" or "web based learning" or "web learning" or "online training" or "e-training" or "Internet training" or "web based training" or "web training" or "massive open online course*" or "MOOC*" or "massively open online course" or "distance education" or "personal learning environment") in the keywords list. In addition, excluding non-article types, 299 articles were retained, and were then reviewed manually according to the content of the articles (including topics and abstracts), excluding duplicates, non-English, literature reviews, and articles not related to AIoL topics. Finally, 224 articles were retained for bibliometric mapping analysis (Figure 1).

Following previous research, we focused on two dimensions of main interest to categorize the 224 articles. The coding scheme was as follows. First, according to the roles of AI mentioned in the collected articles, the four role types of AI were coded as follows: intelligent tutoring systems, profiling and prediction, assessment and evaluation, and adaptive systems and personalization. Next, on the basis of researchers' frameworks (Chen et al., 2020), adopted AI algorithms were classified into 13 types: (a) Bayesian inferencing and networks, (b) evolutionary algorithms, (c) search and optimization, (d) fuzzy set theory, (e) knowledge elicitation methods via interviewing domain experts, (f) neural networks, (g) case-based reasoning, (h) natural language parsing, (i) ontology, (j) data mining, (k) statistical learning, (l) traditional machine learning approaches (including item response theory, linear regression, polynomial regression, Iterative Dichotomiser 3, support vector machine, classification, and clustering), and (m) mixed. Coding was performed by two researchers who read and classified the papers according to the coding scheme, and the coding results of the two researchers showed high consistency (kappa value = 0.89; Lavrakas, 2008).

Figure 1

Article Selection Process for Bibliometric Mapping Analysis and Content Analysis



Data Analysis

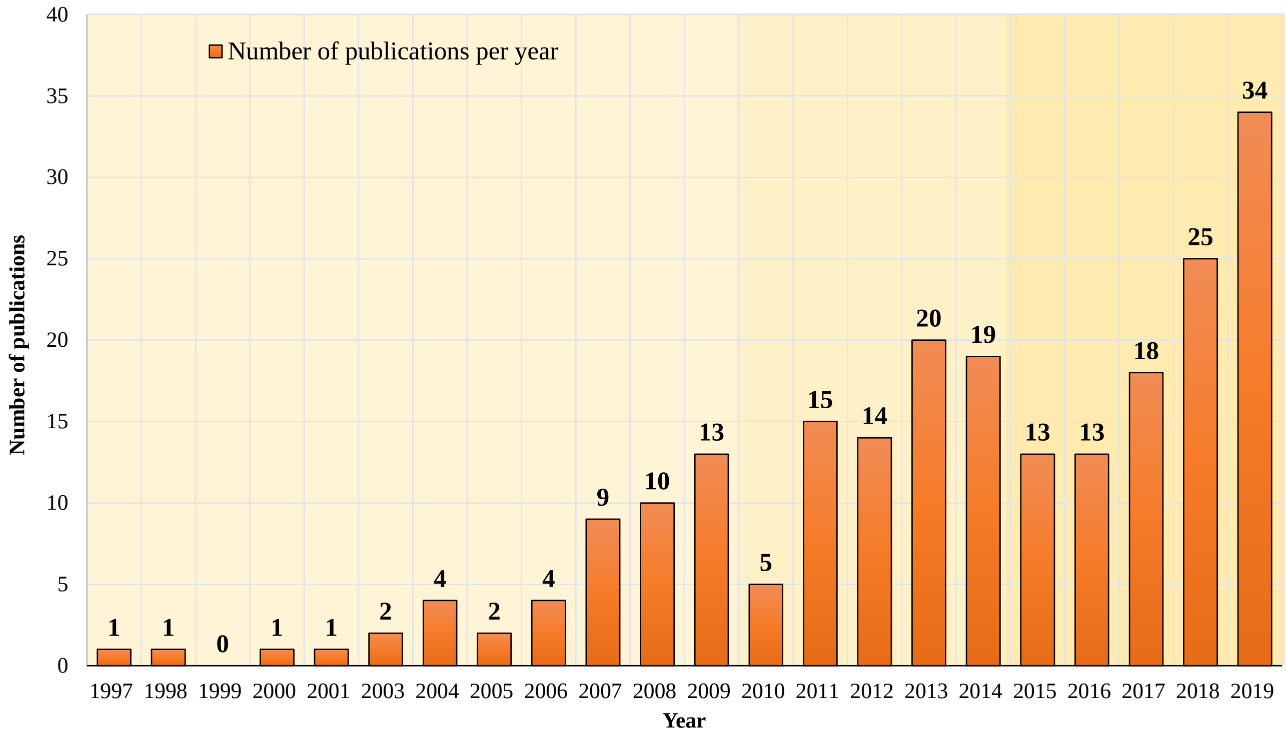
In this study, bibliometric mapping analysis was conducted by employing the VOSviewer software version 1.6.16 to identify the most frequently adopted keywords in AI-supported online learning studies as well as visualizing the citation and co-citation analysis results. The collected data were reviewed by three researchers who examined the descriptive statistics of the data, and discussed and interpreted the findings.

Data Distribution

Figure 2 shows the AIoL studies published from 1997 to 2019. Taking into account the fluctuations in technology, the published AIoL articles were categorized into three time periods, that is, 1997–2009, 2010–2014, and 2015–2019, based on the suggestions of Zawacki-Richter et al. (2019) and Zheng et al. (2016). It was found that there was an increasing trend in AIoL research from the first period (1997–2009, publications = 48) to the second period (2010–2014, publications = 73), and then to the most recent 5 years (2015–2019, publications = 103).

Figure 2

Number of Published Articles on AI in Online Learning Environments from 1997–2019



Note. N = 224.

Results

Analysis of Publication Trends and Country Distribution

Following previous review studies (Hwang & Tsai, 2001), we used the first author's affiliation(s) as the measure to identify the country of origin at the time the article was published. Researchers have suggested that the first author playing the role of the main contributor in a research collaboration is a well-accepted practice in scientific publications, including AI-supported online learning research. Note that for the few first authors in this study who had two or more affiliations in different countries, the main affiliation was manually checked and counted. As shown in Table 2, the most productive countries (top 18) in the field were Taiwan, China, Spain, and the United States.

Table 2

Number of Articles and Rankings of the Most Productive Countries of the AI-Related Publications in the Context of Online Learning Environments

Rank	Country	Total articles, *n*	Articles by year, *n*		
			1997–2009	2010–2014	2015–2019
1	Taiwan	36	13	18	5
2	China	31	3	4	24
3	Spain	25	0	13	12
4	USA	20	6	7	7
5	Greece	11	6	3	2
6	UK	9	3	2	4
7	Turkey	8	0	6	2
8	Italy	7	1	3	3
9	Argentina	6	5	0	1
9	Iran	6	2	2	2
9	Lithuania	6	1	3	2
9	India	6	0	1	5
9	Saudi Arabia	6	0	2	4
14	Australia	4	1	2	1
14	Netherlands	4	0	1	3
16	Brazil	3	1	0	2
16	Germany	3	0	0	3
16	Pakistan	3	0	0	3
-	Top 18	194	42	67	85
-	Other countries (23)	30	6	6	18
-	Total (41)	224	48	73	103

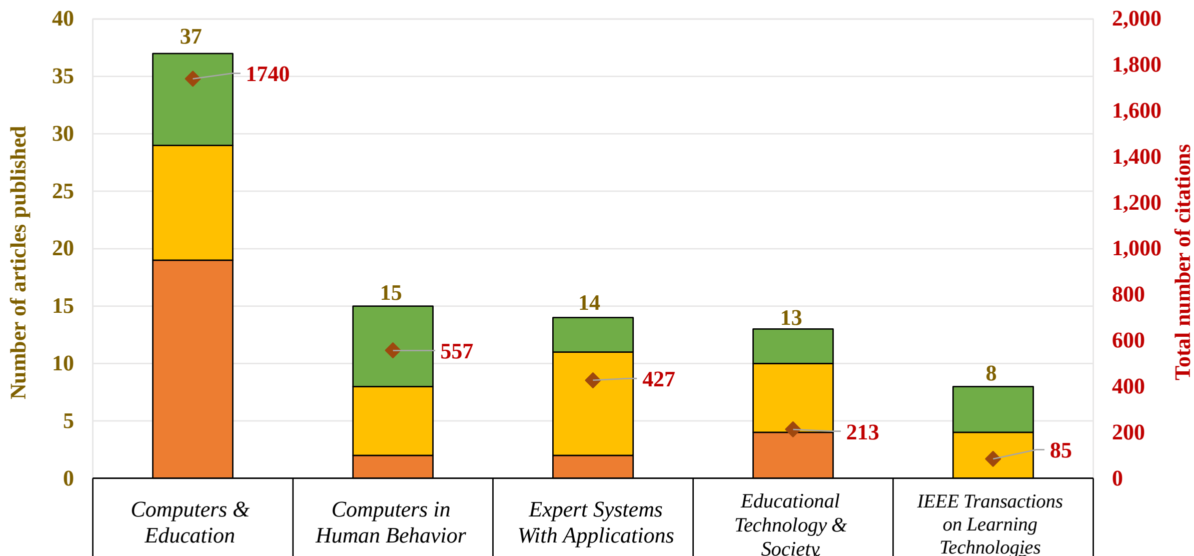
Note. *N* = 224.

Main Journals, Papers, and Authors

In this study, the 224 articles retrieved were published in 84 different journals. Figure 3 shows the five journals with the largest number of articles on AIoL published between 1997 and 2019. They are (a) *Computers & Education*, (b) *Computers in Human Behavior*, (c) *Expert Systems With Applications*, (d) *Educational Technology & Society* and (e) *IEEE Transactions on Learning Technologies*. In Figure 3, it is shown that *IEEE Transactions on Learning Technologies* began to publish AIoL articles only in 2010–2014. This is because the journal was founded in 2008, and a total of four articles were published in that time period. Figure 3 also indicates that the top five journals published 38.84% of AIoL articles. The publications in these five journals are generally highly cited, showing the potential of AIoL research.

Figure 3

Top Five Journals by Total Number of Publications from 1997 to 2019



	Computers & Education	*Computers in Human Behavior*	*Expert Systems With Applications*	*Educational Technology & Society*	*IEEE Transactions on Learning Technologies*
2015-2019	8	7	3	3	4
2010-2014	10	6	9	6	4
1997-2009	19	2	2	4	0
Total # of Citations	1740	557	427	213	85

Note. Only journals with 8 or more publications are included.

Furthermore, co-citation analysis and citation sources were also selected for the journal analysis. In this study, co-citation analysis and citation sources were chosen for creating a map of the most cited journals. The minimum number of citations from sources was adjusted to 20, and the number of sources to be selected was automatically displayed as 37. The created map is presented in Figure 4. It shows that the most-cited journals were *Computers & Education* (citations = 454), *Expert Systems with Applications* (citations = 244), *Lecture Notes in Computer Science* (citations = 180), *Educational Technology & Society* (citations = 103) and *Computers in Human Behavior* (citations = 94).

Figure 4

Most Cited Journals (Co-Citation Analysis)

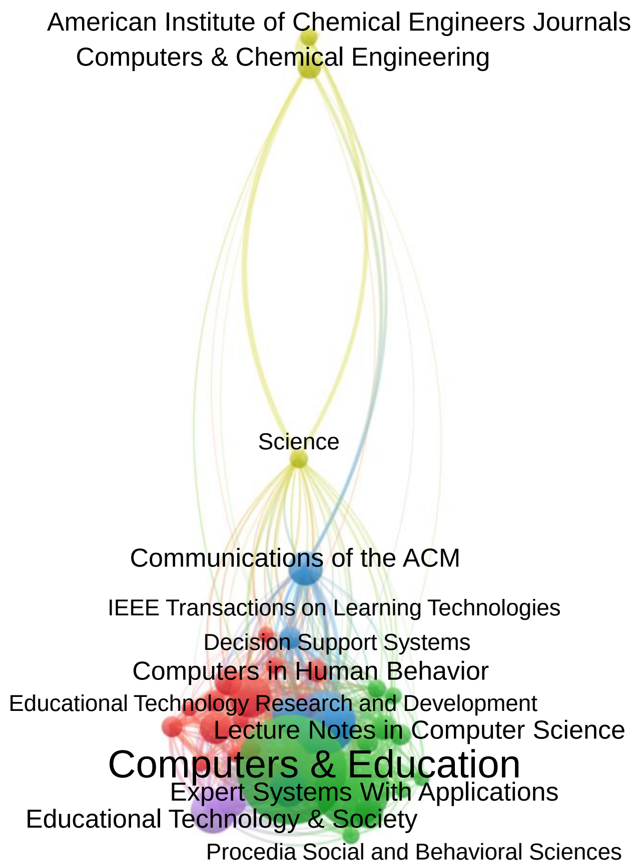


Table 3 shows the five most-cited AIoL articles published between 1997 and 2019. Among the most frequently cited articles, many researchers focused on developing online-learning systems with personalized learning mechanisms to assist online learning, and on adaptively providing learning paths to facilitate individual learners' learning performance. For example, Chen et al. (2005) proposed a personalized e-learning system based on item response theory (PEL-IRT), which considers course material difficulty and learner ability to provide personalized learning paths. García et al. (2007) adopted Bayesian networks to improve the precision of assessing students' learning styles. At the same time, they pointed out that one of the most desirable features of a Web-based education system is that it is adaptable and personalized and can adjust the curriculum or provide assistance to students according to their needs.

Table 3

Five Most-Cited Papers

Rank	Title	Journal	Year	Citations, *n*
1	Personalized e-learning system using item response theory	*Computers & Education*	2005	257
2	Evaluating Bayesian networks' precision for detecting students' learning styles	*Computers & Education*	2007	206
3	Sentiment analysis in Facebook and its application to e-learning	*Computers in Human Behavior*	2014	190
4	Intelligent web-based learning system with personalized learning path guidance	*Computers & Education*	2008	166
5	The politeness effect: Pedagogical agents and learning outcomes	*International Journal of Human-Computer Studies*	2008	147

This study further analyzed those authors who had published five or more papers. In addition, considering the author's influence, we selected authors who had been cited more than 50 times. As shown in Table 4, the most productive authors are Chih-Ming Chen, followed by Maria Virvou, and Analia Amandi. They are clearly active authors in the AIoL field.

Furthermore, co-citation analysis and cited authors were also selected. The minimum number of an author's citations was set to 20 and the number of authors to be selected was automatically set at eight. The created map is shown in Figure 5. It illustrates that Peter Brusilovsky (citations = 74), Chih-Ming Chen (citations = 63), and Cristobal Romero (citations = 41) are the most cited (co-citation) authors in this field.

Table 4

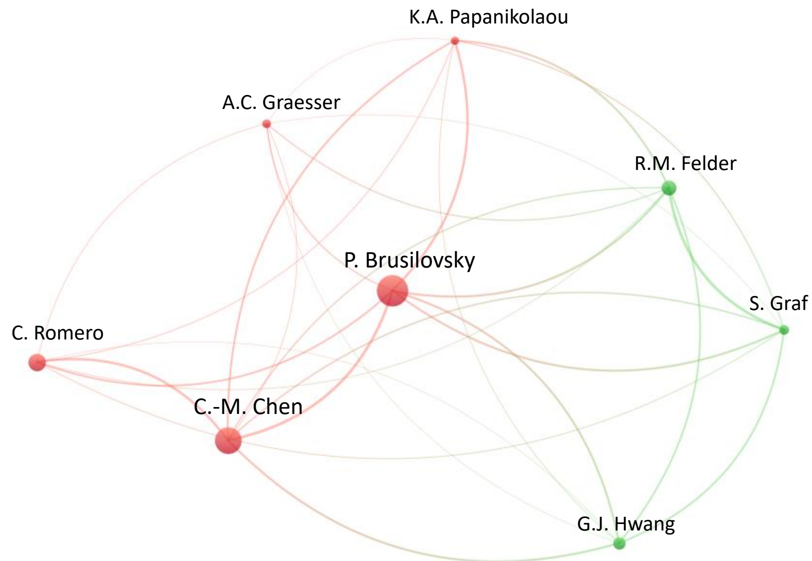
Top Three Authors Ranked by Number of Publications

Author	Country	Publications, *n*	Citations, *n*
Chih-Ming Chen	Taiwan	8	709
Maria Virvou	Greece	6	164
Analia Amandi	Argentina	5	444

Note. Only authors with five or more publications are included in the table.

Figure 5

Most Cited Authors (Co-Citation Analysis)



Main Research Areas

Figure 6 presents the 10 most popular research areas, which predominantly fall into the fields of the social sciences and technology. It can also be seen in Figure 6 that most of the AIoL publications from 1997 to 2019 are in the research areas of education and educational research; others are in computer science and interdisciplinary applications, computer science and artificial intelligence, and engineering, electrical and electronic, etc.

Figure 6

Top 10 Research Areas by Total Number of Publications, 1997–2019

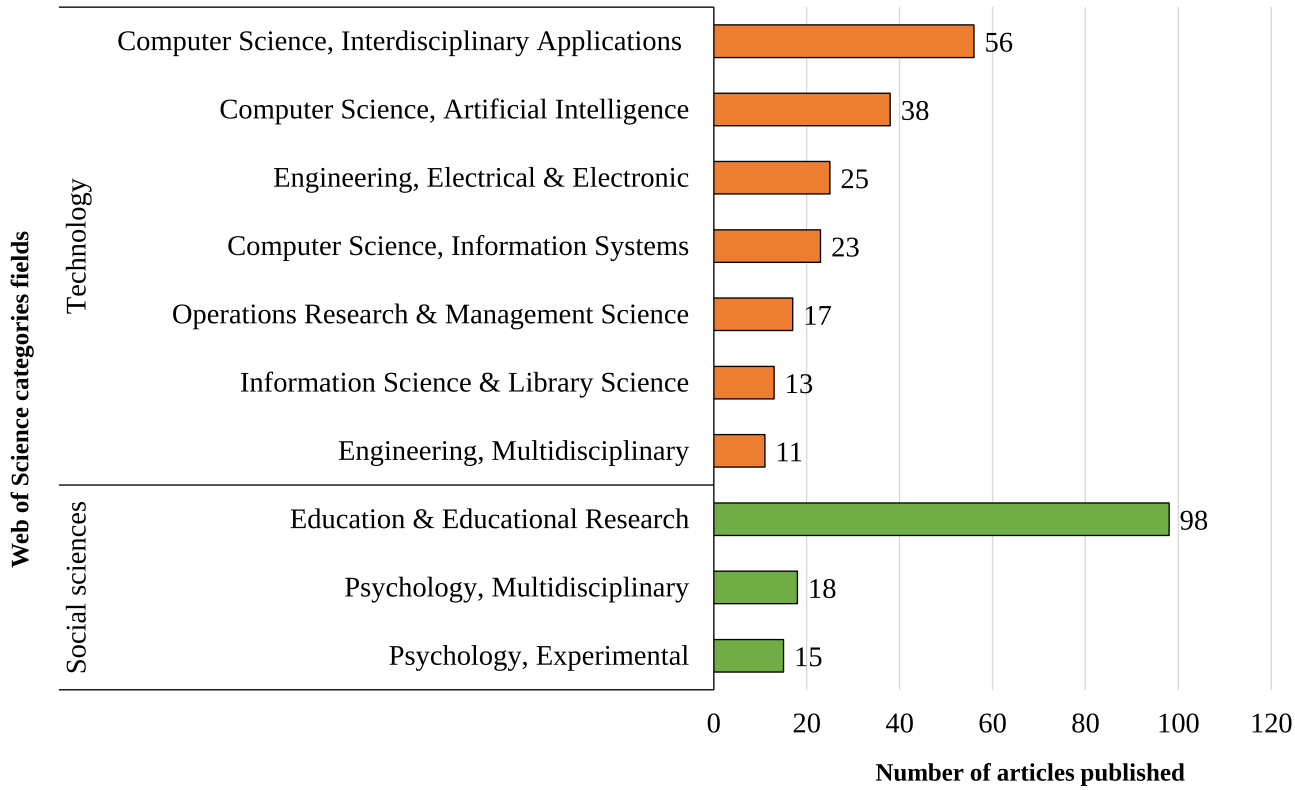
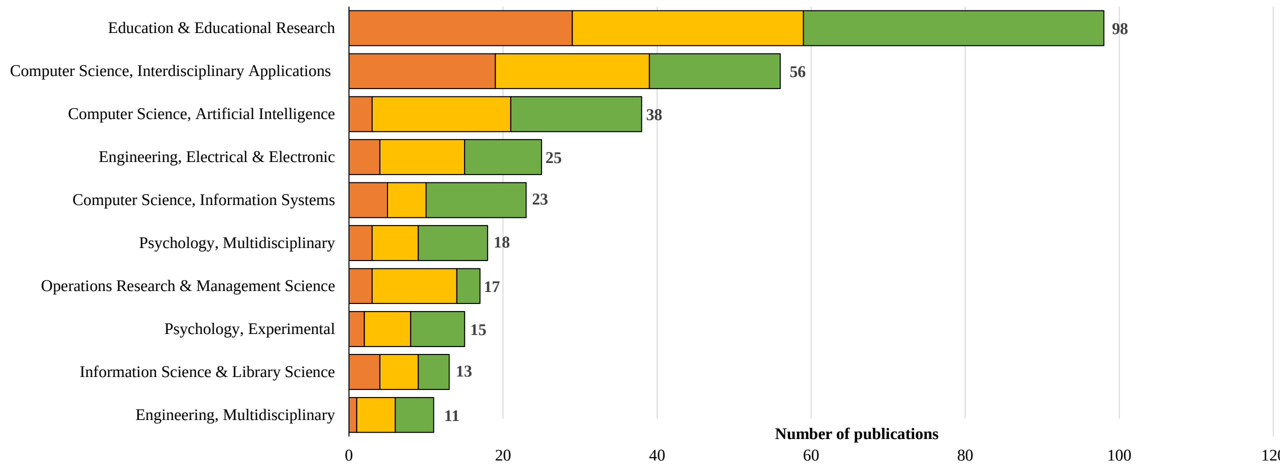


Figure 7 shows the distribution of the top 10 research areas of AIoL. The first AIoL research in 1997 came from two research areas, either education and educational research or computer science and interdisciplinary applications. In 1997–2009, AIoL was most frequently applied to education and educational research, followed by computer science and interdisciplinary applications. In 2010–2014, AIoL was applied most in education and educational research, in the order of computer science, interdisciplinary applications, computer science, artificial intelligence, engineering, electrical and electronic and operations research and management science. In 2015–2019, AIoL was used most in education and educational research, and then in computer science, interdisciplinary applications, engineering, electrical and electronic, computer science, artificial intelligence and computer science, information systems. From the above, it can be seen that in the second period, researchers from different fields paid more attention to the research issues related to AIoL.

Figure 7

Published Literature in the Top 10 Research Areas by Total Number of Publications, 1997–2019



	Engineering, Multidisciplinary	Information Science & Library Science	Psychology, Experimental	Operations Research & Management Science	Psychology, Multidisciplinary	Computer Science, Information Systems	Engineering, Electrical & Electronic	Computer Science, Artificial Intelligence	Computer Science, Interdisciplinary Applications	Education & Educational Research
1997-2009	1	4	2	3	3	5	4	3	19	29
2010-2014	5	5	6	11	6	5	11	18	20	30
2015-2019	5	4	7	3	9	13	10	17	17	39

In the top 10 research areas related to AIoL research, researchers tended to focus on somewhat different concerns. Figure 8 shows the top keywords and newer keywords for each research area. For example, the studies in education and educational research in 1997-2019 focus on how AI technologies were used in developing an interactive online environment to provide personalized supports to individual learners. This is revealed in the top 10 keywords, such as intelligent tutoring systems, e-learning, and interactive learning environments. Moreover, in recent five years (i.e., 2015–2019), the top 10 keywords in this research area show that researchers focused more on the roles of AI in online learning, such as dropout, personalization, and simulations.

Besides, it is found that those with technological background (e.g., Computer Science) are the major researchers who apply AI technologies in interactive online learning environments. It is also found that the development of AI-based online learning systems was the main focus of these studies. Moreover, the roles and the adoption of AI technologies in online learning are diverse (e.g., information extraction, simulations, e-learning tools, educational data mining, neural networks, decision support tools, and recommendation systems). With advances in technology and society's development, some researchers have identified the skills needed to solve problems as one of the significant challenges in distance education.

Figure 8

Top Keywords and Newer Keywords Adopted in AIoL Research

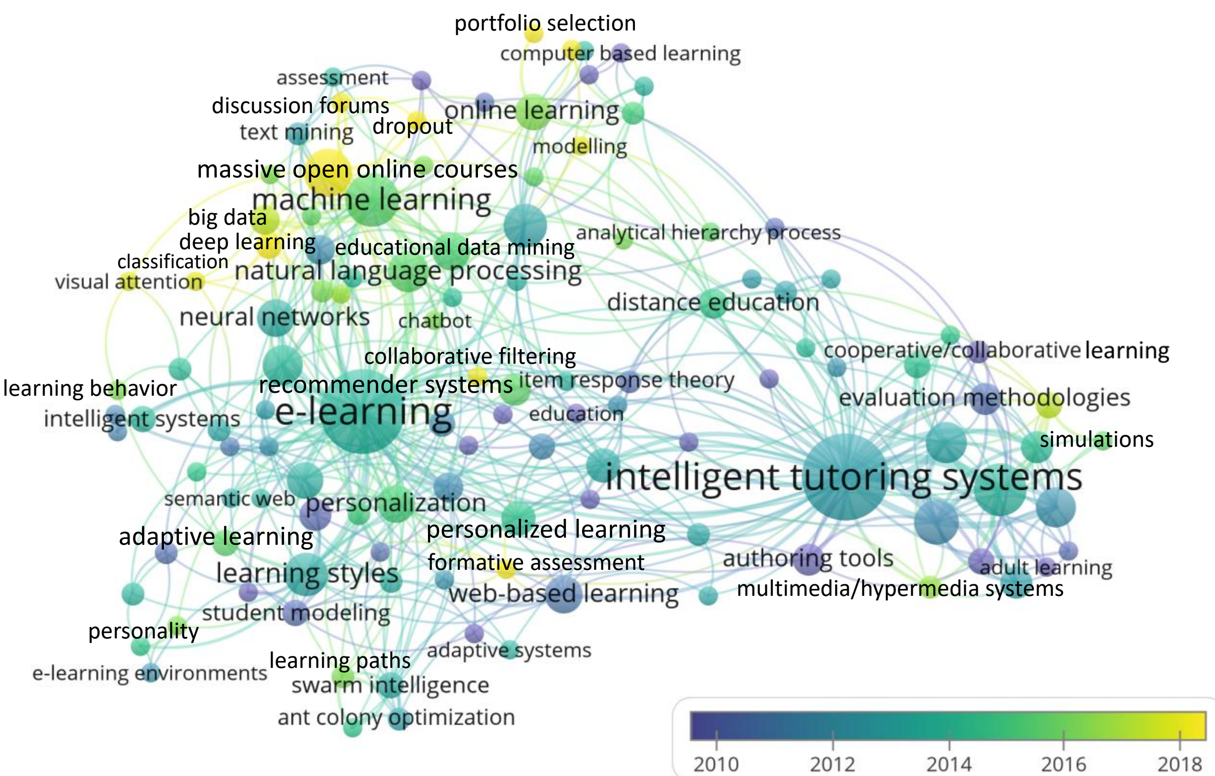
Research areas	Top 10 keywords	New keywords in 2015-2019
Education & Educational Research	intelligent tutoring systems, e-learning, interactive learning environments, distance education and telelearning, teaching/learning strategies, learning styles, massive open online courses, machine learning, human-computer interface and web-based learning.	dropout, discussion forums, ontology, personalization, simulations, e-learning tools, applications in subject areas, adaptive user interface, algorithms competition, artificial intelligence, behavior mining, deep learning and pedagogical issues, etc.
Computer Science, Interdisciplinary Applications	intelligent tutoring systems, interactive learning environments, e-learning, distance education and telelearning, teaching/learning strategies, human-computer interface, architectures for educational technology systems, machine learning, educational data mining and multimedia/hypermedia systems.	information extraction, e-learning tools, applications in subject areas, pedagogical issues, big data, data-driven optimization, deep learning, massive open online courses, simulations, adaptive and intelligent tutoring system, concept map, educational data mining, personalization of learning paths and adaptive systems, etc.
Computer Science, Artificial Intelligence	e-learning, machine learning, data mining, intelligent tutoring systems, interactive learning environments, neural networks, educational data mining, information extraction, decision support tools, recommendation systems.	support vector machines, simulations, online learning, misconception detection and identification, learning strategies, inference system, distance education and information processing, etc.
Engineering, Electrical & Electronic	e-learning, intelligent tutoring systems, artificial intelligence, data mining, online learning, recommendation systems, machine learning, interactive learning environments, individual differences and individualized e-learning.	collaborative filtering, learning styles, personalized learning and recommendation algorithm, etc.
Computer Science, Information Systems	e-learning, adaptive e-learning, intelligent agents, machine learning, neural networks, personalization and virtual learning environments.	collaborative filtering, Felder and Silverman learning style model, learning styles, personalized learning, rating prediction, recommendation algorithm and recommendation systems, etc.
Psychology, Multidisciplinary	e-learning, intelligent tutoring systems, adaptive learning, ant colony optimization, authoring tools, collaborative learning, learners' behavior, learning paths, pedagogical agents and social networks.	artificial intelligence, chatbot, e-health, collaborative learning, human-computer, decision-making, motivation, self-guided intervention and self-management, etc.
Operations Research & Management Science	e-learning, intelligent tutoring systems, data mining, adaptive learning, collaborative learning, individual differences, individualized e-learning, interactive learning environments, item response theory and machine learning.	adaptively, inference system, computational intelligence, distance education and telelearning, etc.
Psychology, Experimental	e-learning, intelligent tutoring systems, adaptive learning, ant colony optimization, authoring tools, collaborative learning, learners' behavior, learning paths, social networks and swarm intelligence.	emotion, personality, user's status, artificial immune system and big data, etc.
Information Science & Library Science	artificial intelligence, computer based learning, databases and learning.	learning styles, personalized learning and recommendation algorithm, etc.
Engineering, Multidisciplinary	e-learning, intelligent tutoring systems and web-based learning.	formative assessment, multi-armed bandit algorithm, upper-confidence bound algorithm, etc.

The Popular Research Topics

The keywords provide essential information about the literature, and some researchers have pointed out that keyword analysis can help researchers clearly understand the trends in specific fields (Guo et al., 2016). In this study, VOSviewer software version 1.6.16 was used to perform a cluster analysis of keywords in order to understand the research issues related to the AIoL field. A total of 657 keywords were used in the AIoL research, and the network maps with keywords occurring with more than two frequencies are shown in Figure 9.

Figure 9

Network Map of Keyword Co-Occurrence on AIoL Research Over Time



To further examine the dynamic change in research topics, the author keywords of AIoL studies in individual time periods were analyzed, as shown in Figure 9 and Table 5, respectively. In the period 1997–2009, AIoL studies focused on the combination of ITS and distance learning; moreover, researchers started to provide support for learners by taking into account their learning status and needs, such as the development of personalized learning systems or ITS (Baylari & Montazer, 2009; Chen, 2009; Chen et al., 2005), algorithms to assess students' learning styles based on their online learning behaviors (García et al., 2007), and models for dropout prediction (Lykourentzou et al., 2009).

In comparison with the publications in 1997–2009, the AIoL studies in 2010–2014 focused more on student-centered learning; that is, using AI technologies (e.g., natural language processing and educational data mining) to enable adaptive learning or personalized learning. For example, Lin et al. (2013) developed a personalized creativity learning system based on the data mining technique of decision trees to provide personalized learning paths for maximizing learner motivation and learning effects, optimizing the learner's creativity performance.

In addition to the provision of online learning supports and the development of ITS, the studies in 2015–2019 focused more on investigating the interactions between learners and learning environments, as well as the provision of interactive learning interfaces in this regard, such as discussion forums, chatbots, educational games, recommendation systems, and decision support tools to facilitate learning outcomes, collaboration, and adaptive learning. For example, several studies aimed to provide personalized learning paths based on individual learners' preferences and/or learning status (e.g., Kurilovas et al., 2015); some studies used AI technologies, such as artificial neural networks (ANNs), to identify learners' problems and provide support or feedback by analyzing their conversation data (Hussain et al., 2019).

In addition, it was found that some keywords, although found in earlier time periods, began to be highlighted in this time. These included terms such as "deep learning," "massive open online courses," "big data," "simulations," "online learning," "learning paths," "experiments," "adaptive e-learning," "recommendation systems," "natural language processing," "machine learning," "educational data mining and personalization," "learning objects," and "neural nets."

Table 5

Frequency of Top Keywords by Year

Keyword	Frequency, *n*			
	1997–2009	2010–2014	2015–2019	Total
intelligent tutoring systems [a]	16	19	22	57
e-learning [a]	13	13	28	54
machine learning [a]	2	6	12	20
interactive learning environments	5	4	10	19
massive open online courses [a]	0	1	14	15
distance education and tele learning	5	3	5	13
artificial intelligence [a]	3	4	5	12
learning styles	4	3	5	12
natural language processing [a]	1	6	5	11
teaching/learning strategies	3	3	5	11
data mining	0	5	5	10
human-computer interface	3	4	3	10

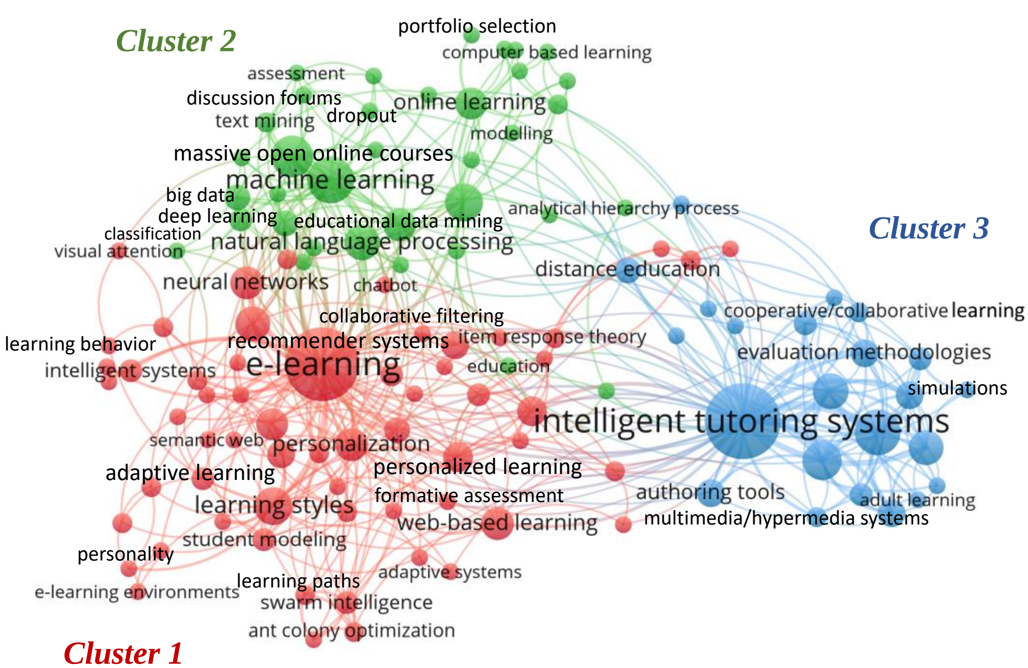
Keyword	Frequency, *n*			
	1997–2009	2010–2014	2015–2019	Total
educational data mining	0	4	5	9
neural networks [a]	2	3	4	9
personalization	2	1	6	9
web-based learning [a]	4	2	3	9
online learning [a]	1	1	6	8
ontology	1	4	3	8
adaptive learning	0	6	1	7
personalized learning	1	3	3	7

Note. Only keywords used ≧ 7 times are included. [a] Initial search terms in this study.

Figure 10 shows that AIoL research themes form three clusters: (a) AI-supported personalized and collaborative online learning, (b) AI-facilitated online-learning management, and (c) development and evaluation of intelligent online-learning systems. These clusters are displayed in three colors: red, green, and blue. There is also a significant correlation between the keywords in each cluster. For example, the red cluster is associated with another keyword, indicating that the research topic is of interest to researchers in the AIoL field and relevant to other areas of AIoL research.

Figure 10

Map of the Structure of Research on AI in Online Learning Environments by Theme



Note. Cluster 1 = AI-supported personalized and collaborative online learning; Cluster 2 = AI-facilitated online-learning management; Cluster 3 = development and evaluation of intelligent online-learning systems.

The main essence of the red cluster (Cluster 1) is AI-supported personalized and collaborative online learning. This is the most important and largest theme cluster in terms of its centrality, overall weight, density, and degree of overlap with the other topics. A large number of its main terms relate to AI, personalization, learners, and/or online learning. The studies in Cluster 1 mainly focused on incorporating AI technologies into online learning to facilitate personalized, adaptive, and collaborative online learning (e.g., Anaya et al., 2013; Chen et al., 2005; Ortigosa et al., 2014; Samarakou et al., 2018). Such a trend has become clearer in recent years, and initiatives include using AI-supported personalized supports to help students complete complex learning tasks (Schiaffino et al., 2008) as well as guide them to make learning plans, examine their own learning status, make reflections, and adjust the plans to promote learning performance (Chen, 2009; Romero et al., 2013; Romero et al., 2019).

Figure 10 also illustrates the strength of the links between the nodes of the keywords in the green cluster (Cluster 2). Thus, these keywords are closely related and form the second theme group in AIoL. The core terms reflect lines of research related to artificial intelligence, online learning, and education. This thematic group focuses on AI-facilitated online-learning management. The outstanding effect of these terms highlights the importance of these concepts in the AIoL framework. The studies in Cluster 2 focused more on using AI technologies to cope with online-learning management problems, such as the provision of automatic evaluation, the prediction of dropout rate or probability, the analysis of learning engagement status using machine learning, natural language processing, educational data mining, support vector machines, and deep learning approaches (e.g., Wise et al., 2017; Xing & Du, 2019; Xing et al., 2019).

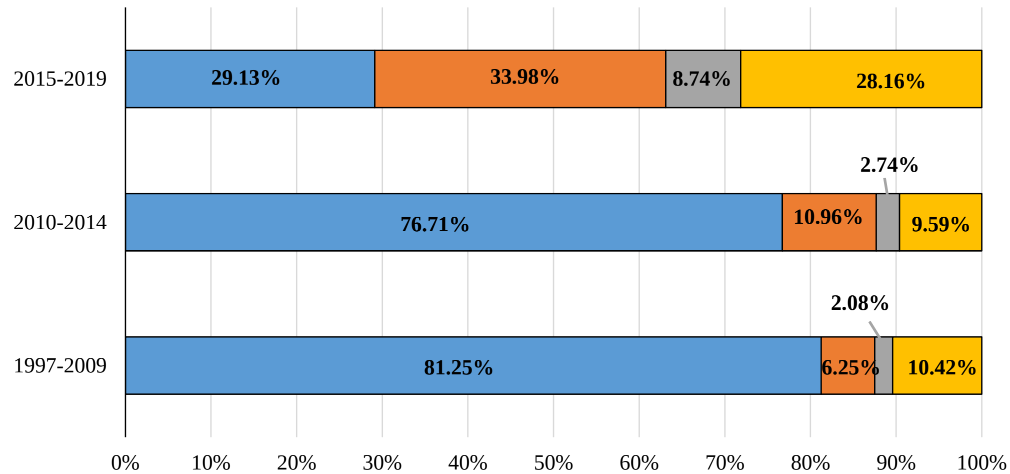
The blue cluster (Cluster 3) focuses on the development and evaluation of intelligent online-learning systems. The core terms of this cluster reflect research interests related to ITS, the development and design of human-computer interfaces and interactive learning environments, and teaching/learning strategies. In these terms, intelligent tutoring systems, interactive learning environments, distance education and telelearning, teaching/learning strategies, and human-computer interface stand out above the rest. This demonstrates the importance of this group's terms throughout the study period. Cluster 3 mainly focuses on the development and evaluation of ITS; for example, Virvou and Alepis (2005) developed an ITS to assist teachers in monitoring students' learning performances by recording and diagnosing individual students' learning processes (e.g., learning logs and test scores).

Roles of AI and Adopted AI Algorithms

Figure 11 shows the roles of AI in the AIoL research. It was found that "intelligent tutoring systems" (55.80%) played the most frequent roles, followed by "profiling and prediction" (20.54%) and "adaptive systems and personalization" (18.30%). Except for "intelligent tutoring systems," the number of AI roles grew from each period.

Figure 11

Roles of AIoL Research in Each Period



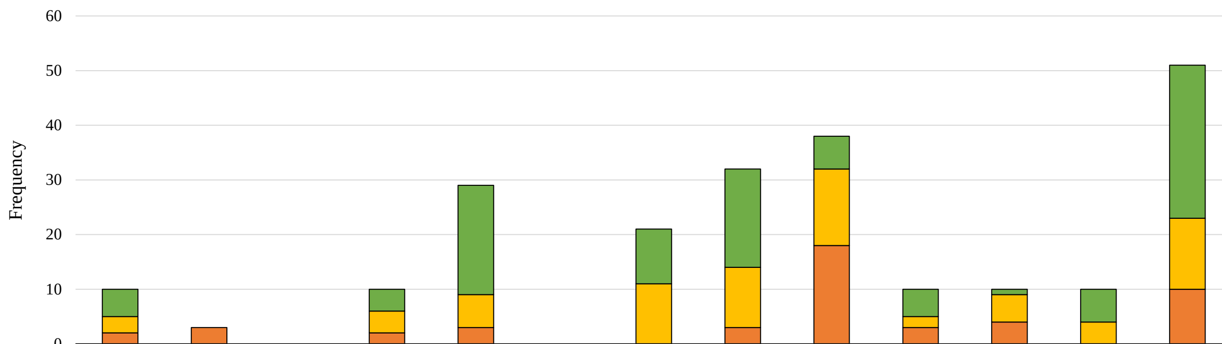
	1997-2009	2010-2014	2015-2019
Intelligent tutoring systems	39	56	30
Profiling and prediction	3	8	35
Assessment and evaluation	1	2	9
Adaptive systems and personalization	5	7	29

Figure 12 shows the adopted AI algorithms. It was found that most studies adopted two or more AI algorithms (22.77%), followed by knowledge elicitation methods via interviewing domain experts (16.96%), ontology (14.29%), neural networks (12.95%), natural language parsing (9.38%), evolutionary algorithms (4.46%), fuzzy set theory (4.46%), data mining (4.46%), statistical learning (4.46%), traditional machine learning approaches (4.46%), and Bayesian inferencing and networks (1.34%). Search and optimization and case-based reasoning were not adopted in any of the studies. The studies which adopted two or more AI algorithms generally aimed to build models to predict students' learning behaviors or performance (e.g., Hussain et al., 2019; Xing et al., 2019). Knowledge elicitation methods via interviewing domain experts were also related to the development of intelligent learning systems for guiding and assessing learners' learning status, providing them with feedback (e.g., de la Peña Esteban et al., 2019). Studies have employed ontology to develop prototypes of e-learning systems and recommender systems as a basis for guiding performance-oriented learning, and to help learners understand through the concepts of ontologies of topics (e.g., Capuano et al., 2012; Romero et al., 2019). From the analysis results, the AIoL research not only focuses on developing and evaluating the effectiveness of ITS or recommendation systems in education, but also highlights the usefulness of AI-related technologies for online learning and teaching. For example, Wang et al. (2018) presented a semantic analysis model to track learners' emotional tendencies, and through emotion quantification and machine learning calculations could predict real-time graduation probability for different learning stages. Xing et al. (2019) adopted an integrated framework of achievement emotions to analyze discussion forum posts to explore the impact of achievement emotions on students' continued participation in MOOCs. Zhou et al. (2019) employed an end-to-end algorithm of deep learning in machine learning to analyze learners' confusion in educational games, and obtained 91.04% precision.

In online learning settings, Scholten et al. (2019) pointed out that chatbots could significantly impact learners during the learning process compared to mere textual guidance.

Figure 12

Adopted Technologies of AIoL Research in Each Period



	Evolutionary algorithms	Bayesian inferencing and networks	Search and optimization	Fuzzy set theory	Neural networks	Case-based reasoning	Natural language parsing	Ontology	Knowledge elicitation methods via interviewing domain experts	Statistical learning	Traditional machine learning approach	Data mining	Mixed
2015-2019	5	0	0	4	20	0	10	18	6	5	1	6	28
2010-2014	3	0	0	4	6	0	11	11	14	2	5	4	13
1997-2009	2	3	0	2	3	0	0	3	18	3	4	0	10

Conclusion, Discussion, and Suggestions

Owing to the progress and popularity of computer and network technologies, online learning environments have significantly changed in the past decades. Building on the work of Shukla et al. (2019), Wong et al. (2019), and Zawacki-Richter et al. (2019), this study analyzed the AIoL journal articles published from 1997 to 2019 using the bibliometric mapping analysis approach by categorizing the publications into three time periods. Based on our analysis, we found five findings merited further discussion.

Productive Countries

The finding related to most productive countries differed slightly from what has been reported in several review studies of AI in education (AIEd); for example, Zawacki-Richter et al. (2019) reported that the most productive countries in AIEd research for higher education were the United States, followed by China and Taiwan. This implies that Asian researchers are more focused on applying AI technologies to online learning environments than to other learning contexts. More importantly, the researchers from these countries/regions (e.g., Taiwan and China) are not native speakers of English. The quality and productive research output could be due to government policies promoting online learning. For example, Taiwan's government conducted a nationwide program from 2003 to 2012 to promote e-learning, which significantly encouraged Taiwan's researchers to focus on e-learning studies (Hwang & Tsai, 2011; Tsai et al., 2010).

AIoL Studies Accepted in Top-Ranking Journals

The largest number of AIoL studies was published in *Computers & Education*, followed by *Computers in Human Behavior* and *Expert Systems with Applications*. These journals are well recognized as top-ranking journals in the fields of educational technology, experimental psychology, and computer science, respectively. This implies that AIoL studies have gained widespread acceptance among scholars in different fields. This is consistent with the analysis results of AIoL research areas.

Government Programs May Promote Research

Based on the productivity and citation analysis, it was found that the most influential author of AIoL research is Chih-Ming Chen, a researcher from Taiwan, who has published 8 articles with 709 citations. This output could be due to the nationwide e-learning promotion program initiated by Taiwan's government, which encourages researchers, more than in other countries/regions, to focus on e-learning studies (Tsai et al., 2010).

Three Clusters of Research Topics

In term of research topics, AIol studies can be categorized into three clusters, that is, "AI-supported personalized and collaborative online-learning," "AI-facilitated online-learning management," and "development and evaluation of intelligent online-learning systems." This finding provides a good reference for researchers hoping to design future AIoL studies.

Roles of AI and Adopted AI Algorithms

"Intelligent tutoring systems" have played the most important role in AIoL, followed by "profiling and prediction," and "adaptive systems with personalization." This is consistent with the findings of previous studies regarding research issues (Hwang et al., 2020; Zawacki-Richter et al., 2019). Most studies adopted two or more AI algorithms to develop models of learners' online learning to predict their learning performance, followed by "knowledge elicitation methods via interviewing domain experts," while search and optimization and case-based reasoning are less commonly employed. From the results of keyword analysis and AI technology adoption surveys, it was found that the application of AI-related algorithms, as well as discussion forums, chatbots, and educational games, showed a significant growth trend (Scholten et al., 2019; Xing et al., 2019; Zhou et al., 2019).

In summary, this study not only reveals the increasing emphasis on AIoL research, but also reports several important trends in this field. Based on the findings of this study, several suggestions for designing AIoL research as well as developing policies to promote AIoL in the future are presented.

1. It is recommended that educational research institutions in various countries follow the practice of some leading countries and regions to promote AIoL through large-scale and long-term programs.

2. In addition to domains in which AIoL research is frequently applied, researchers are encouraged to apply the AIoL approach to less often studied domains such as architecture, economics and management, as well as cross-disciplinary applications.

3. It could also be interesting to extend the AIoL approach to blended learning contexts or ubiquitous learning contexts to investigate the impacts of this approach on integrated real-world and virtual-world learning.

4. One important and often overlooked research topic is the use of AI technologies to optimize learners' online learning experiences (e.g., information literacy, curiosity, emotions, etc.) to increase their learning engagement and reduce dropout rates.

5. Further research is needed into the role that AIoL could play in providing personalized learning guidance and supports as well as learning paths.

6. It is crucial to investigate students' higher order thinking and behavioral patterns, and not only their learning achievements in AIoL contexts.

References

Anaya, A. R., Luque, M., & García-Saiz, T. (2013). Recommender system in collaborative learning environment using an influence diagram. *Expert Systems with Applications, 40*(18), 7193–7202. https://doi.org/10.1016/j.eswa.2013.07.030

Baylari, A., & Montazer, G. A. (2009). Design a personalized e-learning system based on item response theory and artificial neural network approach. *Expert Systems with Applications, 36*(4), 8013–8021. https://doi.org/10.1016/j.eswa.2008.10.080

Capuano, N., Mangione, G. R., Pierri, A., & Salerno, S. (2012). Learning goals recommendation for self regulated learning. *International Journal of Engineering Education, 28*(6), 1373–1379. https://www.researchgate.net/profile/Nicola-Capuano/publication/281116341_Learning_Goals_Recommendation_for_Self_Regulated_Learning/links/55d701e308ae9d65948c204e/Learning-Goals-Recommendation-for-Self-Regulated-Learning.pdf

Carin, L. (2020). On artificial intelligence and deep learning within medical education. *Academic Medicine, 95*(11S), S10-S11. https://doi.org/10.1097/ACM.0000000000003630

Chen, C.-M. (2009). Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance. *Expert Systems with Applications, 36*(5), 8816–8829. https://doi.org/10.1016/j.eswa.2008.11.026

Chen, C.-M., & Lain, Y.-Y. (2020). Developing a computer-mediated communication competence forecasting model based on learning behavior features. *Computers & Education: Artificial Intelligence, 1*, Article 100004. https://doi.org/10.1016/j.caeai.2020.100004

Chen, C.-M., Lee, H.-M., & Chen, Y.-H. (2005). Personalized e-learning system using item response theory. *Computers & Education, 44*(3), 237–255. https://doi.org/10.1016/j.compedu.2004.01.006

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers & Education: Artificial Intelligence, 1*, Article 100002. https://doi.org/10.1016/j.caeai.2020.100002

de la Peña Esteban, F. D., Lara Torralbo, J. A., Lizcano Casas, D., & Martínez Rey, M. A. (2019). Expert system for problem solving in distance university education: The successful case of the subject "operations management". *Expert Systems, 36*(5), Article e12444. https://doi.org/10.1111/exsy.12444

García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education, 49*(3), 794–808. https://doi.org/10.1016/j.compedu.2005.11.017

Guo, L., Xu, F., Feng, Z., & Zhang, G. (2016). A bibliometric analysis of oyster research from 1991 to 2014. *Aquaculture International, 24*(1), 327–344. https://doi.org/10.1007/s10499-015-9928-1

Han, E.-R., Yeo, S., Kim, M.-J., Lee, Y. H., Park, K.-H., & Roh, H. (2019). Medical education trends for future physicians in the era of advanced technology and artificial intelligence: An integrative review. *BMC Medical Education, 19*(1), 1–15. https://doi.org/10.1186/s12909-019-1891-5

Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review, 52*(1), 381–407. https://doi.org/10.1007/s10462-018-9620-8

Hwang, G.-J., & Tsai, C.-C.. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 42*(4), E65-E70. https://doi.org/10.1111/j.1467-8535.2011.01183.x

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers & Education: Artificial Intelligence, 1*, Article 100001. https://doi.org/10.1016/j.caeai.2020.100001

Khanal, S. S., Prasad, P. W. C., Alsadoon, A., & Maag, A. (2020). A systematic review: Machine learning based recommendation systems for e-learning. *Education and Information Technologies, 25,* 2635–2664. https://doi.org/10.1007/s10639-019-10063-9

Kurilovas, E., Zilinskiene, I., & Dagiene, V. (2015). Recommending suitable learning paths according to learners' preferences: Experimental research results. *Computers in Human Behavior, 51*, 945–951. https://doi.org/10.1016/j.chb.2014.10.027

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage.

Lee, D., Watson, S. L., & Watson, W. R. (2020). The influence of successful MOOC learners' self-regulated learning strategies, self-efficacy, and task value on their perceived effectiveness of a massive open online course. *The International Review of Research in Open and Distributed Learning, 21*(3), 81–98. https://doi.org/10.19173/irrodl.v21i3.4642

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development, 59*(5), 593–618. https://doi.org/10.1007/s11423-010-9177-y

Lin, C. F., Yeh, Y.-C., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education, 68*, 199–210. https://doi.org/10.1016/j.compedu.2013.05.009

Lin, P.-H., Wooders, A., Wang, J. T.-Y., & Yuan, W. M. (2018). Artificial intelligence, the missing piece of online education? *IEEE Engineering Management Review*, *46*(3), 25–28. https://doi.org/10.1109/EMR.2018.2868068

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education, 53*(3), 950–965. https://doi.org/10.1016/j.compedu.2009.05.010

Martin, F., Sun, T., & Westine, C. D. (2020). A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & Education, 159*, Article 104009. https://doi.org/10.1016/j.compedu.2020.104009

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior, 31*, 527–541. https://doi.org/10.1016/j.chb.2013.05.024

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education, 21*(1), 135–146. https://doi.org/10.1002/cae.20456

Romero, L., Saucedo, C., Caliusco, M. L., & Gutiérrez, M. (2019). Supporting self-regulated learning and personalization using ePortfolios: A semantic approach based on learning paths. *International Journal of Educational Technology in Higher Education, 16*(1), Article 16. https://doi.org/10.1186/s41239-019-0146-1

Samarakou, M., Tsaganou, G., & Papadakis, A. (2018). An e-learning system for extracting text comprehension and learning style characteristics. *Educational Technology & Society, 21*(1), 126–136. https://www.jstor.org/stable/26273874

Schiaffino, S., Garcia, P., & Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students. *Computers & Education, 51*(4), 1744–1754. https://doi.org/10.1016/j.compedu.2008.05.008

Scholten, M. R., Kelders, S. M., & Van Gemert-Pijnen, J. E. W. C. (2019). An empirical study of a pedagogical agent as an adjunct to an eHealth self-management intervention: What modalities does it need to successfully support and motivate users?. *Frontiers in Psychology*, *10*, 1063. https://doi.org/10.3389/fpsyg.2019.01063

Shukla, A. K., Janmaijaya, M., Abraham, A., & Muhuri, P. K. (2019). Engineering applications of artificial intelligence: A bibliometric analysis of 30 years (1988–2018). *Engineering Applications of Artificial Intelligence*, *85*, 517–532. https://doi.org/10.1016/j.engappai.2019.06.010

Tang, K.-Y., Chang, C. Y., & Hwang, G.-J. (2021). Trends in artificial intelligence-supported e-learning: A systematic review and co-citation network analysis (1998–2019). *Interactive Learning Environments*, 1–19. https://doi.org/10.1080/10494820.2021.1875001

Tsai, C.-C., Chen, N.-S., & Chen, G.-D. (2010). The current status and future of e-learning in Taiwan. *Innovations in Education and Teaching International, 47*(1), 5–7. https://doi.org/10.1080/14703290903525879

Virvou, M., & Alepis, E. (2005). Mobile educational features in authoring tools for personalised tutoring. *Computers & Education, 44*(1), 53–68. https://doi.org/10.1016/j.compedu.2003.12.020

Wang, L., Hu, G., & Zhou, T. (2018). Semantic analysis of learners' emotional tendencies on online MOOC education. *Sustainability*, *10*(6), Article 1921. https://doi.org/10.3390/su10061921

Wise, A. F., Cui, Y., Jin, W., & Vytasek, J. (2017). Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling. *The Internet and Higher Education, 32*, 11–28. https://doi.org/10.1016/j.iheduc.2016.08.001

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G.-J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human–Computer Interaction, 35*(4–5), 356–373. https://doi.org/10.1080/10447318.2018.1543084

Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research, 57*(3), 547–570. https://doi.org/10.1177/0735633118757015

Xing, W., Tang, H., & Pei, B. (2019). Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. *The Internet and Higher Education, 43*, Article 100690. https://doi.org/10.1016/j.iheduc.2019.100690

Yu, H., Miao, C., Leung, C., & White, T. J. (2017). Towards AI-powered personalization in MOOC learning. *npj Science of Learning*, *2*(1), Article 15. https://doi.org/10.1038/s41539-017-0016-3

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, 39. https://doi.org/10.1186/s41239-019-0171-0

Zheng, P., Liang, X., Huang, G., & Liu, X. (2016). Mapping the field of communication technology research in Asia: Content analysis and text mining of SSCI journal articles 1995–2014. *Asian Journal of Communication, 26*(6), 511–531. https://doi.org/10.1080/01292986.2016.1231210

Zhou, Y., Xu, T., Li, S., & Shi, R. (2019). Beyond engagement: An EEG-based methodology for assessing user's confusion in an educational game. *Universal Access in the Information Society*, *18*(3), 551–563. https://doi.org/10.1007/s10209-019-00678-7

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 23, Number 1

February – 2022

Using Educational Data Mining Techniques to Identify Profiles in Self-Regulated Learning: An Empirical Evaluation

Eric Araka[1], Robert Oboko[2], Elizaphan Maina[1], and Rhoda Gitonga[1]
[1]Kenyatta University, [2]University of Nairobi

Abstract

With the increased emphasis on the benefits of self-regulated learning (SRL), it is important to make use of the huge amounts of educational data generated from online learning environments to identify the appropriate educational data mining (EDM) techniques that can help explore and understand online learners' behavioral patterns. Understanding learner behaviors helps us gain more insights into the right types of interventions that can be offered to online learners who currently receive limited support from instructors as compared to their counterparts in traditional face-to-face classrooms. In view of this, our study first identified an optimal EDM algorithm by empirically evaluating the potential of three clustering algorithms (*expectation-maximization*, *agglomerative hierarchical*, and *k-means*) to identify SRL profiles using trace data collected from the Open University of the UK. Results revealed that agglomerative hierarchical was the optimal algorithm, with four clusters. From the four clusters, four SRL profiles were identified: poor self-regulators, intermediate self-regulators, good self-regulators, and exemplary self-regulators. Second, through correlation analysis, our study established that there is a significant relationship between the SRL profiles and students' final results. Based on our findings, we recommend agglomerative hierarchical as the optimal algorithm to identify SRL profiles in online learning environments. Furthermore, these profiles could provide insights on how to design a learning management system which could promote SRL, based on learner behaviors.

Keywords: educational data mining, EDM, self-regulated learning, SRL profile, algorithm, agglomerative hierarchical clustering, clustering algorithm

Introduction

The increased adoption of technology to enhance learning along a continuum that ranges from physical classrooms to online learning has opened valuable opportunities for decision makers in institutions of learning. The current coronavirus pandemic has also forced many institutions of higher learning to adopt online teaching and learning resulting in many new datasets being generated. These datasets can be used to understand how to enhance learning pedagogies such as self-regulated learning (SRL) (Coman et al., 2020). Machine learning offers the potential to explore educational data to detect learner profiles that can be used to provide targeted interventions to online students. The behavior of students in online learning environments can be measured from log data that contains page views, access to learning materials, frequency and duration of logins, assignment submission deadlines, number of clicks on learning materials, number of forum posts by students, and quiz and assignment scores (Aljohani et al., 2019; Alshabandar et al., 2018; Barnard et al., 2010; Kuzilek et al., 2017; Lodge & Corrin, 2017).

Over the last three decades since the recognition of SRL, there has been emphasis on the importance of SRL skills in relation to academic achievement. SRL is a process through which students manage their learning while being guided by their own motivation, behavior, and metacognition. Students with high levels of SRL skills are able to play an active role in achieving their academic goals (Klug et al., 2011; Pintrich, 2004). Learners who employ SRL strategies such as time management, help-seeking, and self-monitoring perform better than those who do not (Broadbent & Poon, 2015). The identification of SRL profiles in online learning has been based mostly on data collected using student self-report tools (Barnard et al., 2010; Broadbent & Fuller-Tyszkiewicz, 2018; Valle et al., 2008; Yot-Domínguez & Marcelo, 2017). These self-report tools include the Online Self-Regulated Learning Questionnaire (Barnard et al., 2010), the Motivated Strategies Learning Questionnaire (Broadbent & Fuller-Tyszkiewicz, 2018; Valle et al., 2008), and the Survey of Self-Regulated Learning with Technology at the University (Yot-Domínguez & Marcelo, 2017). Although self-report tools are easy to implement when measuring SRL, students tend to overestimate their skills, and hence may fail to capture the actual learning behaviors exhibited during an online course (Araka et al., 2020; Gašević et al., 2017). Learners also often may fail to recall the strategies they use during learning as self-report tools are employed before or after the learning process (Broadbent & Fuller-Tyszkiewicz, 2018; Elsayed et al., 2019). Literature reveals that researchers rely on both trace data collected from educational systems such as learning management systems (LMSs) and self-report data (Ainscough et al., 2019; Çebi & Güyer, 2020; Gašević et al., 2017; Kim et al., 2018). Using trace data to measure SRL strategies has been viewed as unobtrusive since the tools are deployed without learners being aware and, therefore, they do not affect leaners' engagement behavior and performance (Schraw, 2010). Educational data mining (EDM) techniques therefore are likely to measure and profile learners more accurately as compared to self-report tools, as they use actual datasets collected from online learning environments. EDM is becoming extremely valuable for educators and decision makers especially in higher education institutions as it provides great opportunities for exploring huge datasets already stored in many learning environments. EDM has made it possible to detect students' online learning behavior (Khanna et al., 2016; Siemens & Baker, 2012; Winne & Baker, 2013). With EDM techniques being part of machine learning algorithms, there is a need for an empirical analysis to establish the optimal values of parameters and the best algorithm to use with educational data.

Recent studies have investigated the measurement and promotion of SRL on massive open online courses (MOOCs) (Kizilcec et al., 2017; Maldonado-Mahauad et al., 2018; Wong et al., 2020). However, there is little evidence to show how university and college students self-regulate when engaging in open and distributed learning using LMSs which are commonly used to facilitate distance learning in higher education (Araka et al., 2020). In view of this, the current study investigates SRL profiles using a dataset collected from the Open University, UK so as to allow for more research on the relationship between students' learning behaviors and academic performance. Moreover, the study seeks to inform researchers, educators, and designers of online learning environments about the optimal EDM techniques that can be used to design and provide targeted interventions for ODL students.

The profiling of learners into groups based on students' SRL skills has been done using step-wise cluster analysis (Ainscough et al., 2019; Çebi & Güyer, 2020; Valle et al., 2008; Yot-Domínguez & Marcelo, 2017), a K-means clustering algorithm (Li et al., 2018), latent class analysis (Barnard et al., 2010), and agglomerative hierarchical clustering algorithm (Gašević et al., 2017). Our review of literature revealed that different data mining techniques vary in their performance depending of the source of the dataset and type of e-learning environment. For example, EDM techniques used to measure and promote SRL for MOOCs are different from those used in LMSs. Moreover, there is a lack of evidence showing which algorithm performs better in identifying SRL profiles from data collected from an LMS. In view of this, the current study explored the appropriate EDM algorithm that could be used to profile online learners and group them into appropriate clusters so as to allow for the provision of interventions geared towards supporting SRL. Specifically, the study was guided by the following research questions:

1. What EDM techniques are currently being used to identify SRL profiles in online learning environments?
2. What EDM algorithm is optimal in identifying SRL profiles in online learning environments?
3. What SRL profiles can be identified from students who engage in online learning?
4. How are the SRL profiles identified from an online learning dataset associated with students' final results?

In this paper, the literature review section discusses previous studies on the profiling of learners according to their SRL strategies. Next, the methodology used to address the research questions is outlined. A review of the current EDM techniques being used to identify SRL profiles follows. Then, experimental evaluation of the EDM algorithms identified from the review is presented. The results section offers the findings of the experimental evaluation. Finally, the conclusions and future implications of the study are discussed.

Literature Review

Current research has proved that data mining techniques can be used to enrich decision making in different domains such as finance, healthcare, and e-commerce by transforming raw data into information (Madni et al., 2017). Educational data mining is also critical in analyzing data to improve pedagogical aspects of teaching and learning (Coman et al., 2020). Open and distributed learning has tremendously grown and been adopted by institutions of higher learning (Saadati et al., 2021). Students who engage in online learning, especially higher education, are supposed to play an active role in the learning process. However, literature reveals that students, individually or collectively, do not regulate their own learning (Cerezo et al., 2016; Dabbagh & Kitsantas, 2005). Additionally, online learners are not directly supported by instructors as compared to their counterparts in traditional face-to-face learning. Consequently, there is a need to provide support for SRL and student engagement that is geared towards enhancing self-regulatory skills (Silvola et al., 2021). In view of this, there is need to examine how learners behave and engage in online learning so as to establish the right interventions to be provided to the learners. Student engagement in online learning, especially behavioral and cognitive aspects, are observable elements that indicate how students participate and get involved in their learning activities (Silvola et al., 2021). SRL, on the other hand, is concerned with learners being proactive in their learning, taking their own initiative to control their learning by setting academic goals and identifying strategies to achieve those goals (Azevedo, 2009; Zimmerman, 1990). In the current study, student level engagement behavior in online learning activities is therefore an indicator of SRL level. For instance, a highly active student, identified through the number of resources accessed and the learning activities engaged on, is in control of the learning process and therefore exhibits a high level of self-regulatory behavior. Students' engagement behaviors and learning patterns in online learning environments, such as an LMS, can be measured using trace data. The dataset features may include content or page views, frequency of logins, access to learning materials, forum posts by students, and quiz and assignment scores (Araka et al., 2020).

Previous studies indicate that distinct profiles of SRL exist among students who engage in online and blended learning. The profiles can be identified using EDM methods applied to self-report data, trace data, or both. For example, Barnard et al. (2010) used latent class analysis to identify five profiles of self-regulators: super self-regulators, competent self-regulators, forethought-endorsing self-regulators, performance/reflection self-regulators, and non/minimal self-regulators. The algorithm was applied on data collected using a self-report online questionnaire known as the Online Self-Regulated Learning Questionnaire (OSLQ) (Barnard et al., 2010).

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

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

Finally, Çebi and Güyer (2020) presented various learning activities to students using the Moodle LMS. The learning activities included tutorials, video, concept maps, exercises, and summary, highlight, and forum activities. The data were collected from three sources: self-report data, trace data, and assessment data. Cluster analysis involving hierarchical clustering and k-means were used to identify three clusters. The study, however, was limited to only three weeks and a single course and, therefore, researchers may not have had the opportunity for proper observance of behavior change in leaners as far as SRL is concerned. Table 1 presents a summary of the rest of the studies that we reviewed.

Table 1

Summary of SRL Profiles Identified From Previous Studies

Reference	SRL profiles identified	Data source	Technique to identify profiles
Valle et al. (2008)	• Intermediate SRL level • High SRL level • Low SRL level	SR	Two-step cluster analysis
Barnard et al. (2010)	• Super self-regulators • Competent self-regulators • Forethought-endorsing self-regulators • Performance/reflection self-regulators • Non/minimal self-regulators	SR	Latent class analysis
Yot-Domínguez and Marcelo (2017)	• High-level regulators • Low-level regulators	SR	Stepwise cluster analysis • Hierarchy analysis • Ward method • K-means analysis
Gašević et al. (2017)	• Formative assessment • Summative assessment through trial and error • Studying reading materials • Video watching with formative assessment	SR and TD	Agglomerative hierarchical clustering (based on Ward's algorithm)
Li et al. (2018)	• Early completers • Late completers • Early dropouts • Late dropouts	TD	K-means clustering
Broadbent and Fuller-Tyszkiewicz (2018)	• Minimal regulators • Restrained regulators • Calm self-reliant capable regulators • Anxious capable collaborators • Super regulators	SR	Latent profile analysis

Kim et al. (2018)	• Self-regulation • Partial self-regulation • Non-self-regulation	SR and TD	K-medoids clustering
Ainscough et al. (2019)	• High self-regulators • Medium self-regulators • Low self-regulators	SR and TD	Two-step cluster analysis
Peach et al. (2019)	• Early birds • On time • Low engagers • Crammers • Sporadic outliers (unclustered learners)	TD	Mathematical framework (based on dynamic time warping kernel and clustering algorithm)
Çebi and Güyer (2020)	• Cluster 1: Students with least interaction • Cluster 2: Intense interaction with video, example, and forum activities • Cluster 3: Students who spend more time on tutorial, exercises, concept map, summary, and highlight activities	SR, TD and AD	Cluster analysis • Hierarchical clustering • K-means clustering

Note. SR = self-report. TD = trace data. AD = assessment data.

In their review, Elsayed et al. (2019) established that among the EDM techniques used in measuring SRL, clustering algorithms were most common (Elsayed et al., 2019). The EDM algorithms used in profiling SRL in online learning environments included expectation-maximization (Bouchet et al., 2013; Manzanares et al., 2017; Matcha et al., 2019), k-means (Çebi & Güyer, 2020; Kizilcec et al., 2013; Li et al., 2018; Valdiviezo et al., 2013; Yot-Domínguez & Marcelo, 2017; Zheng et al., 2020), agglomerative hierarchical (Cicchinelli et al., 2018; Maldonado-Mahauad et al., 2018; Matcha et al., 2019; Sun et al., 2016), and process and sequence mining (Kinnebrew et al., 2013; Matcha et al., 2019; Rodriguez et al., 2014; Wong et al., 2019). Classification algorithms included k-nearest neighbor (Syuhada et al., 2020), neural networks (Yu et al., 2018), and logistic regression (Bosch et al., 2018). The review of literature reveals a lack of evidence concerning which algorithm performs better in identifying SRL profiles from trace data collected from online learning environments. Consequently, the current study explores which EDM algorithm would be best to profile learners, group them into appropriate clusters, and establish the association between profiles and students' final results.

Methodology

To address the research questions in the current study, we used a mixed method approach. First, a systematic review of the literature on current EDM techniques used to profile SRL was carried out. The review followed five steps of systematic review methodology (Khan et al., 2003). The review stages included (a) framing the research questions, (b) identifying relevant literature, (c) setting the articles' assessment criteria, (d) presenting review results, and (e) discussing the results. This review formed the foundation for the second study which involved experimental evaluation of EDM algorithms in order to establish the optimal algorithm to identify SRL profiles from a dataset obtained from the Open University in the UK.

Finally, correlation analysis was used to identify the association between the SRL profiles and students' academic performance.

Review of Educational Data Mining Techniques Used in Profiling SRL

The reviewed articles in this study were iteratively searched from international journals and databases which included Google Scholar, SCOPUS, Science Direct, Elsevier, ERIC, IEEE Xplore, and ACM digital libraries. The articles were searched using keywords: "educational data mining techniques" AND "learner analytics" AND "measurement of self-regulated learning" AND "assessment of self-regulated learning" AND "clickstream data" AND "student behaviors" AND "online learning" AND "self-regulated learning profiles." A total of 72 papers was identified. After reading the full text of each article and applying the inclusion criteria described in Khan et al., 2003, 48 papers were removed. A total of 24 papers, 12 journal articles and 12 conference articles, met the inclusion criteria. A summary is presented in Table 2.

Inclusion Criteria

There were four inclusion criteria used to obtain relevant literature for the systematic review:

a) articles that used EDM or LA techniques for measuring SRL in online learning environments;

b) articles that described machine learning experiments using trace data obtained from higher institutions of learning;

c) articles that described experiments using self-report data integrated with trace data to construct models for measuring SRL; and

d) articles that described software application(s) that implemented EDM algorithm(s) for SRL measurement.

Systematic Review Results

In this section, we present a review of the literature on current EDM techniques used to group learners into various SRL profiles according to their behavioral interactions in online learning environments. Table 2 presents a summary.

Table 2

Algorithms Used to Measure SRL in Online Learning Environments

Reference	Data source	Feature set	EDM technique	Algorithm used
Bouchet et al. (2013)	MetaTut or trace data & self-	• Page views • Page visits • Note-taking duration • Session duration	Clustering	Expectation-maximization

	report data	• Assessment scores • No. of quizzes completed		
Zheng et al. (2020)	Trace data	• Structural views • Functional shows • Design additions/edits • Note taking	Clustering	K-means
Valdiviezo et al. (2013)	LMS trace data	• Course hits • Course views • Assignment views • Forum events • Resources views • Message events • Quiz events	Clustering	K-means
Maldonado-Mahauad et al. (2018)	MOOC trace data & self-report data	• Video views • Video reviews • Assessment trials • Course completion status • Assessment reviews • Assessment passes	Clustering	Agglomerative hierarchical
Manzanares et al. (2017)	LMS trace data & self-report data	• Access to course materials • Access to assessments • Access to teacher feedback • Forum participation • Mean access rates per day	Clustering	Expectation-maximization
Cicchinelli et al. (2018)	LMS trace data	• View content indices • View course organization • View exercises • Solve quizzes • View content	Clustering	Agglomerative hierarchical
Kizilcec et al. (2013)	MOOC trace data	• Forum activity • In-video assessments • Demographic features	Clustering	K-means
Park et al. (2018)	LMS trace data	• Video clicks • Quiz submissions • Assignment submissions	Clustering	Probability model based clustering (Poisson mixture model)
Sun et al. (2016)	LMS trace data & self-report data	• Number of assessment attempts • Assessment scores • Time spent of each online lecture • Lecture completion status	Clustering	Agglomerative hierarchical
Matcha et al. (2019)	Trace data	• Videos with multiple-choice questions (MCQs) • Reading materials with MCQs • Exercises	Clustering & temporal data mining	Agglomerative hierarchical & expectation-maximization, process & sequence mining
Rodriguez et al. (2014)	PLE trace data	• Blogs • Video annotations • Bookmarks	Temporal data mining	Process mining

		• Tags • Comments • Excerpts		
Wong et al. (2019)	MOOC trace data	• Video views • Quizzes • Assignments • Forum discussions	Temporal data mining	Sequential pattern mining using equivalence classes
Kinnebrew et al. (2013)	Betty's Brain system trace data	• Reading • Editing • Querying • Explaining • Quizzing	Temporal data mining	Differential sequence mining
Cerezo et al. (2020)	LMS trace data	• Forum discussion • Quiz • Resources views • URL views • Course performance	Temporal data mining	Inductive miner
Yu et al. (2018)	LMS trace data	• Video navigations • Assignment views • Quiz views • Discussion sessions	Temporal data mining	Neural networks (LSTM, RNN, & GRU)
Di Mitri et al. (2016)	Multimodal data	• Heart rate • Step count • Weather condition • Learning activity	Classification	Regression analysis
Bosch et al. (2018)	LMS trace data	• No. of weeks logged in • Total logins • No. of events per login • Total interaction events • Access to materials • Grade views • Quiz attempts • Correct quiz answers • Exam attempts • Correct exam attempts • Forum post views • Forum posts created	Classification	Logistic regression
Syuhada et al. (2020)	Trace data	• Features not mentioned	Classification	K-nearest neighbor
Trevors et al. (2016)	Multimodal data & self-report data	• Eye tracking patterns • Study tools • Metacognitive ratings	Statistical modeling	Correlation analysis
Montgomery et al. (2019)	LMS trace data	• Access location • Access time (of the day) • Online login frequency • Online login regularity • Quiz review pattern • Course materials views	Statistical modeling	Association & correlational analysis

Jansen et al. (2020)	MOOC trace data & self-report data	• Video interaction events • Quiz interaction events • Marking reading as completed • Submission of assignment • Page navigations • Visits & posts on forums	Statistical modeling	Statistical modeling
Jo et al. (2016)	LMS trace data & self-report data	• Login frequency • Login regularity • Total login time	Statistical modeling	Statistical modeling
Rodriguez et al. (2019)	LMS trace data & self-report data	• Video clicks • Slide clicks	Statistical modeling	Binomial regression
Crossley et al. (2016)	MOOC trace data	• Video interaction • Forum interaction • Page views • Assignments	natural language processing (NLP) tools	WAT, TAALES, TAACO, ReaderBench, & SEANCE

The EDM algorithms identified from the review can be categorized into clustering algorithms, temporal data mining, and other techniques that include natural language processing (NLP) and classification. These EDM categories are discussed in this section.

Clustering Algorithms

Clustering algorithms represent the class of unsupervised machine learning techniques that classify learners into groups based on the similar interaction behaviors inferred from log data. Several clustering algorithms have been identified in this study including expectation-maximization, K-means and agglomerative hierarchical.

Expectation-maximization (EM) has been used to identify SRL behaviors and profile learners into various groups based on interaction behaviors. For example, Bouchet et al. (2013) used EM to identify three clusters of learners from trace data derived from learner behaviors. Similarly, Manzanares et al. (2017) used EM to group learners into three clusters. Since the EM algorithm involves predetermining the number of clusters, Manzanares et al. (2017) used the bi-stage cluster node to determine the value of k. Additionally, Matcha et al. (2019) investigated how EM can cluster students based on learning sequences which were also used to identify the SRL strategies based on the indicators learners used. The agglomerative hierarchical was utilized to identify learning patterns from the SRL strategies identified from the clusters (Matcha et al., 2019). In this study, various learning behaviors were identified: reading-oriented students, exercise-oriented students, and students oriented toward MCQs and video. Other students exhibited diverse behaviors such as the use of exercises, video views, and MCQs in learning. Three groups of learners were identified: high-, moderate-, and low-level SRL engagers.

The K-means clustering algorithm was used in a number of studies. The K-means algorithm iteratively divides a given dataset into a number of distinct number of clusters. The value of *k* therefore represents the number of dissimilar clusters identified from a dataset. The data points in each cluster are similar to each other and dissimilar from data points in other clusters (Nuankaew et al., 2019). In their study, Zheng et al. (2020) employed the K-means clustering algorithm to identify profiles in SRL for learners taking STEM courses in engineering design. In this study, *principle component analysis* was used to reduce the high-dimensionality of the data (Zheng et al., 2020). Given that K-means is an unsupervised machine learning algorithm, the number of clusters needed to be pre-determined; the ball statistic was used to establish the optimal number of clusters. The clusters identified in that study included competent self-regulated learners, minimally self-regulated learners, cognitive-oriented self-regulated learners, and reflective self-regulated learners. However, the study had limitations. For one, the indicators of the SRL were based on an *Energy 3D* learning environment that is specifically used by engineering students. The study therefore may not be applicable across other non-engineering courses and programs. Similarly, Valdiviezo et al. (2013) used the k-means algorithm to identify three clusters: high, medium, and minimal access and usage levels, based on students' online interaction behaviors from virtual learning interaction (VLI) data from the Moodle LMS. The highest level of self-regulated learners, according to the study, were those students who had the greatest amount of interaction on forums, in terms of responding, viewing and adding discussions, quizzes, reading and writing messages, and accessing online learning resources. The k-means gives accurate results for similar experiments in the area of modelling student learning behaviors (Valdiviezo et al., 2013). Finally, Kizilcec et al. (2013) used k-means to identify groups of learners based on engagement behaviors as measured from trace data collected on a MOOC platform.

The agglomerative hierarchical algorithm, which helps to identify an unknown number of clusters given variables of interest from datasets, was also identified in the review. For example, Sun et al. (2016) investigated the effect of SRL on performance trajectory in a flipped classroom using the agglomerative hierarchical clustering algorithm. Six trajectory groups based on students' performance and trace data from interactions on the LMS were identified. The agglomerative hierarchical algorithm has also been used in other studies to identify distinct groups of learners based on their SRL variables as reported using an MSLQ self-report tool (Pardo et al., 2016, 2017). The groups were then used to investigate the association between the students' online activity interactions and academic performance. Additionally, agglomerative hierarchical, based on Ward's method, was used to identify profiles of learners from trace data (Cicchinelli et al., 2018).

Temporal Data Mining

Temporal data mining encompasses two main techniques: process mining and sequence mining. A process mining algorithm is used to describe the paths followed by learners in an online learning environment (Rodriguez et al., 2014). Sequential mining on the other hand is used to identify sequences of learning activities using learner interaction logs. The objective is to determine the path followed by online students and the frequency of the activities carried out by the students (Wong et al., 2019). Sequence mining and process mining have been used to identify learning paths especially on MOOC platforms. Process mining is usually carried out before sequence mining. This helps generate process models that are based on students' time-stamped actions captured during the learning process. The sequence of learning actions that students perform during a learning episode will help understand the path followed by learners. The output is exploited for cluster analysis (Matcha et al., 2019).

In the review, several studies used process and sequence mining to investigate the presence of SRL strategies detected in trace data from both MOOCs and LMSs. For example, Cerezo et al. (2020) used process mining to measure SRL process from students' interaction data generated from the Moodle LMS. The inductive miner algorithm was used to produce process models that demonstrated students' learning behaviors. The process models reproduced students' interaction on the LMS. In that study, the highly regulated students were found to have followed the learning paths suggested by the instructor. This group of learners also performed activities related to forum discussions. In a related study, Kinnebrew et al. (2013) used differential sequence mining to identify and classify learners into groups based on their behaviors. Sequence mining requires that the trace data, which contain student interaction logs that indicate students' learning patterns, is first transformed into a sequence of actions. In this study, sequence mining was used to identify frequent patterns from a set of sequences. The indicators captured by Betty's Brain, a software agent, included *read*, *edit*, *query*, *explain*, and *quiz*. The algorithm analyses the sequence of actions and classifies learners into three groups: high, low, and medium engagers. Likewise, Maldonado-Mahauad et al. (2018), in their study whose main objective was to identify learning interaction sequences, clustered students with similar behavioral characteristics. Process mining was used to first identify the learning paths followed by learners in a MOOC course. The interaction sequences that were used for exploratory analysis were later used for clustering of learners into profiles. For clustering, agglomerative hierarchical was used to cluster learners according to the interaction sequences they followed. Three groups were identified: sampling learners (low level SRL), as well as comprehensive learners and targeting learners, who exhibited similar SRL behaviors.

Other EDM Techniques

Other machine learning algorithms and statistical modeling were also applied on multimodal data to measure the SRL of online learners (Di Mitri et al., 2016, 2017; Trevors et al., 2016). Likewise, statistical modeling, such as association techniques, along with other techniques, such as confirmatory factor analysis, was applied. For example, Crossley et al., 2016 used natural language processing (NLP) tools to complement trace data with language properties in understanding learner behavior especially from forum posts. The indices of NLP that were used included text length, social collaboration, sentiment analysis, text cohesion, syntactic complexity, lexical sophistication, and quality of writing. Classification techniques have also been used to categorize learners according to their learning patterns. For example, logistic regression was used to classify learners into different demographic and underrepresented groups based on trace data collected from an LMS (Bosch et al., 2018). Statistical modeling and frequency of learning activities were also performed so as to better understand various online learning behaviors. For example, Jansen et al. (2020) investigated the levels of compliance to the SRL interventions that were provided to learners by the MOOC. Neural network techniques have also been used to determine the extent to which students' learning paths conform to the pre-determined course structure. The page clickstream data was used, including the sequence of video interactions, assignment and quiz navigations, welcome page views, and discussion sessions (Yu et al., 2018).

Sources of Data and Feature Sets for Measuring SRL

As presented in Table 2, the sources of datasets and the features sets used for profiling learners based on behavior patterns were also investigated. A majority of the studies used trace data collected from LMSs such as Moodle (Cerezo et al., 2020; Jo et al., 2016; Manzanares et al., 2017; Montgomery et al., 2019; Sun et al., 2016; Valdiviezo et al., 2013), and Canvas (Park et al., 2018; Rodriguez et al., 2019). In measuring SRL and identifying SRL profiles, some studies relied on trace data in MOOCs such as those offered at the Coursera website (Crossley et al., 2016; Jansen et al., 2020; Kizilcec et al., 2013; Maldonado-Mahauad et al., 2018; Wong et al., 2019). Other online learning environments included Energy 3D (Zheng et al., 2020), Betty's Brain (Kinnebrew et al., 2013), and LON-CAPA (Bosch et al., 2018). Moreover, datasets collected from agent-based software applications such as MetaTutor, an agent-based system purposely developed to promote SRL, were used to profile and cluster learning according to students' interaction behaviors in a virtual learning environment (VLE; Bouchet et al., 2013).

The findings reveal that the dataset features used for profiling and measuring SRL in online learning are determined by the type of e-learning environment from which the data was collected. For example, for studies that used LMS data, the indicators include forum-related activities such as posting and updating forums, viewing, and replying to other students' posts. Other learning activities considered are quiz events such as quiz completion status and submission time in relation to the set deadlines, course module views, writing and reading messages, and the frequency and regularity of student logins (Jo et al., 2016; Montgomery et al., 2019; Valdiviezo et al., 2013). For trace data from MOOCs, learning activities related to video interactions such as video views and reviews, quiz events, assignment attempts and reviews, and course completion status were considered (Jansen et al., 2020; Kizilcec et al., 2013; Maldonado-Mahauad et al., 2018; Wong et al., 2019). Some researchers used multimodal data to measure SRL (Di Mitri et al., 2016, 2017; Trevors et al., 2016).

Discussion on the Systematic Review

The main objective of the systematic review was to identify the EDM techniques that are currently being used to measure SRL using trace data from online learning environments. The results reveal that clustering algorithms are more commonly used as compared to temporal data mining and classification algorithms. Our findings agree with the results obtained from a previous review (Elsayed et al., 2019). The study also revealed that the EDM algorithms currently being used in measuring and profiling SRL in online learning environments include expectation-maximization (Bouchet et al., 2013; Manzanares et al., 2017; Matcha et al., 2019), k-means (Çebi & Güyer, 2020; Kizilcec et al., 2013; Li et al., 2018; Valdiviezo et al., 2013; Yot-Domínguez & Marcelo, 2017; Zheng et al., 2020), agglomerative hierarchical (Cicchinelli et al., 2018; Maldonado-Mahauad et al., 2018; Matcha et al., 2019; Sun et al., 2016), and process mining (Kinnebrew et al., 2013; Matcha et al., 2019; Rodriguez et al., 2014; Wong et al., 2019). Classification algorithms that have been used in the reviewed studies include k-nearest neighbor (Syuhada et al., 2020), neural networks (Yu et al., 2018) and logistic regression (Bosch et al., 2018).

From the review, it can be established that SRL dataset features from online learning environments could potentially be influencing the type of algorithm used to profile learners based on their SRL skills. For example, it can be observed that process and sequence mining were mostly applied on datasets collected from MOOCs and PLEs where the feature sets considered were the video interaction events, quiz, and assignment type and submissions timelines (Kinnebrew et al., 2013; Matcha et al., 2019; Rodriguez et al., 2014; Wong et al., 2019). On the other hand, clustering algorithms were mostly applied on LMS data where the feature sets such as module and page views, login frequency and regularity, and assignment and quiz views and scores were mostly considered (Cicchinelli et al., 2018; Jo et al., 2016; Manzanares et al., 2017; Montgomery et al., 2019; Park et al., 2018; Sun et al., 2016; Valdiviezo et al., 2013).

From the review, it can be argued that there is no empirical evidence that shows which EDM algorithm for profiling SRL using online learning datasets is optimal. The experimental evaluation carried out in the next section was therefore conducted with the objective of establishing the optimal EDM algorithm for profiling learners according to their course interaction behaviors.

Experimental Evaluation

In this section, we describe the experiment carried out to compare the clustering algorithms identified from the systematic review. The algorithms identified from the literature review were compared to determine the optimal number of clusters formed by the best performing algorithm. For research questions two and three, a dataset collected from a virtual learning environment at the Open University in the UK was applied to the algorithms identified to profile learners into clusters and also test for any association between SRL profiles and academic performance.

Dataset Description and Preprocessing

The dataset collected from the Open University in the UK was used to identify the optimal clustering algorithm and the optimal number of clusters in online learning. The Open University Learning Analytics Dataset (OULAD) was chosen for this study as it represents students' actual behaviors in an online LMS as compared to other sets of data (Jha et al., 2019). The dataset contains three categories of student information: demographic, interactions in the form of logs, and assessments. The dataset is organized in tabular form with seven files. The data represents 22 courses and 32,593 students, their assessment results, and their interactions with a virtual learning environment (VLE) (Kuzilek et al., 2017). The current study used the dataset extracted from the *studentInfo*, *vle*, and *studentVle* tables (N = 735). The dataset represents students' interactions in one course offered in two semesters. The interactions are represented by the number of clicks/visits to specific learning resources and activities, such as course notes in the form of HTML pages and pdf files, and learning activities in the form of discussion forums and quizzes (Kuzilek et al., 2015). According to Kuzilek et al. (2017), the resources that were being accessed by the students included the course homepage, external and internal URLs, course subpages, resources, discussion forums, a glossary, collaboration tools, and course content. Table 3 presents a summary of the OULAD dataset and its features (Kuzilek et al., 2017).

Table 3

Summary of the Open University Learning Analytics Dataset

Table name	Records, *n*	Description	Table attributes
courses	22	Information about the courses	code_module, code_presentation, module_presentation_length
studentInfo	32,593	Demographic information about the students	code_module, code_presentation, id_student, gender, region, highest_education, imd_band, age_band, num_of_prev_attempts, studied_credits, disability, final_result
studentRegistration	32,593	Registration of the student for a course presentation	code_module, code_presentation, id_student, date_registration, date_unregistration
assessments	196	Assessments for every course presentation	code_module, code_presentation, id_assessment, assessment_type, date, weight
studentAssessments	173,740	Assessments submitted by the students	id_assessment, id_student, date_submitted, is_banked, score
vle	6,365	Online learning resources and materials	id_site, code_module, code_presentation, activity_type, week_from, week_to
studentVle	1,048,575	Student interaction with the VLE resources	code_module, code_presentation, id_student, id_site, date, sum_click

After feature extraction, which was done using *id_ student*, *code_module* and *code_presentation* as unique identifiers from three files that included *studentVle*, *studentInfo* and *courses*, one file was generated containing 5 columns and 735 rows. The extracted file contained one course named AAA, which was offered in two separate semesters to two separate cohorts one in 2013 and another in 2014 represented by 2013J and 2014J. Table 4 presents a summary of the sample dataset obtained for experimental evaluation. The sum of clicks captured students' interactions with various resources stored on the VLE. The clickstream data, which is also referred as number/sum of clicks in this study, represents the number of interactions students made when accessing various learning activities and resources.

Table 4

Summary of the Preprocessed Sample OULAD Dataset for Module AAA

Semester	Student ID	Sum of clicks	Final results
2013J	100893	744	Pass
2014J	258587	6,609	Distinction
2014J	2606802	306	Fail
2013J	101781	4,104	Pass
2013J	129955	1,011	Withdrawn
2013J	102806	1,944	Pass
2013J	146188	597	Fail
2013J	102952	1,150	Pass
2013J	147793	155	Withdrawn
2014J	263251	2,485	Pass
2013J	1035023	1,896	Pass
2014J	268733	3	Fail

The preprocessed data was then imported to a Python environment where various clusters were formed using the three algorithms: k-means, expectation-maximization, and agglomerative hierarchical. The algorithms were implemented for clustering and visualization in the RStudio environment where the statistical evaluations were computed.

Experimental Procedure

First, the Python programming language was used to visualize scatterplots for the clusters formed by the three algorithms being compared, where the number of clusters was varied from 3 to 10 for each algorithm. Secondly, the clusters formed were compared using internal validation indices provided by the clValid (Brock et al., 2008) and the NbClust (Charrad et al., 2014) R packages. The functions were used to compare the algorithms based on the internal information of the data by evaluating the “goodness” and quality of the clusters formed. The outputs from the evaluations were used to determine the optimal number of clusters and the best performing algorithm (Rodriguez et al., 2019; Van-Craenendonck & Blockeel, 2015). The clValid uses the Dunn index, Connectivity, and the Silhouette index to establish the optimal number of clusters and the best performing algorithms (Brock et al., 2008). The NbClust, on the other hand, determines the optimal number of clusters in the dataset using the results of 30 inbuilt indices (Charrad et al., 2014).

Experimental Evaluation Results

In this section, experimental results for the three clustering algorithms are discussed. First, we examine the results of the three clustering algorithms. Second, the clustering evaluation carried out to determine the most appropriate algorithm with the optimal number of clusters is described. Last, we present the results of the test for independency between the optimal clusters and students' final academic achievement.

As presented in Figures 1, 2, and 3, the scatterplots demonstrate the clusters formed by the K-means, expectation-maximization, and agglomerative hierarchical algorithms while varying the number of clusters from 3 to 10.

Figure 1

Clustering Using the K-Means Algorithm

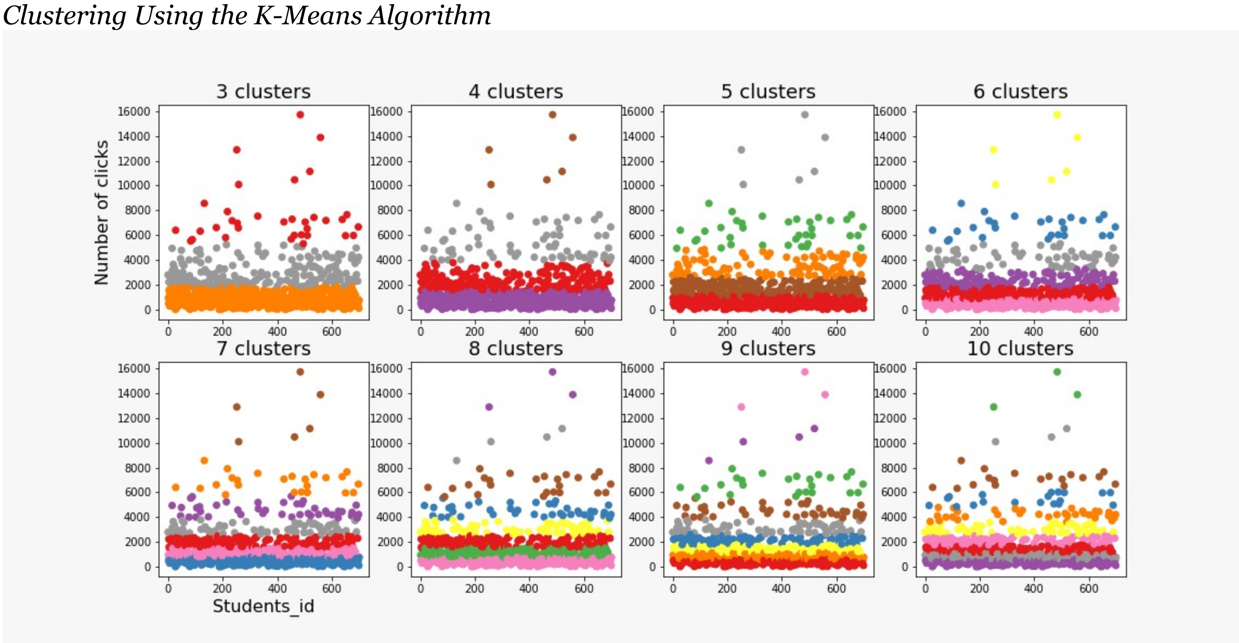


Figure 2

Clustering Using the Expectation Maximization Algorithm

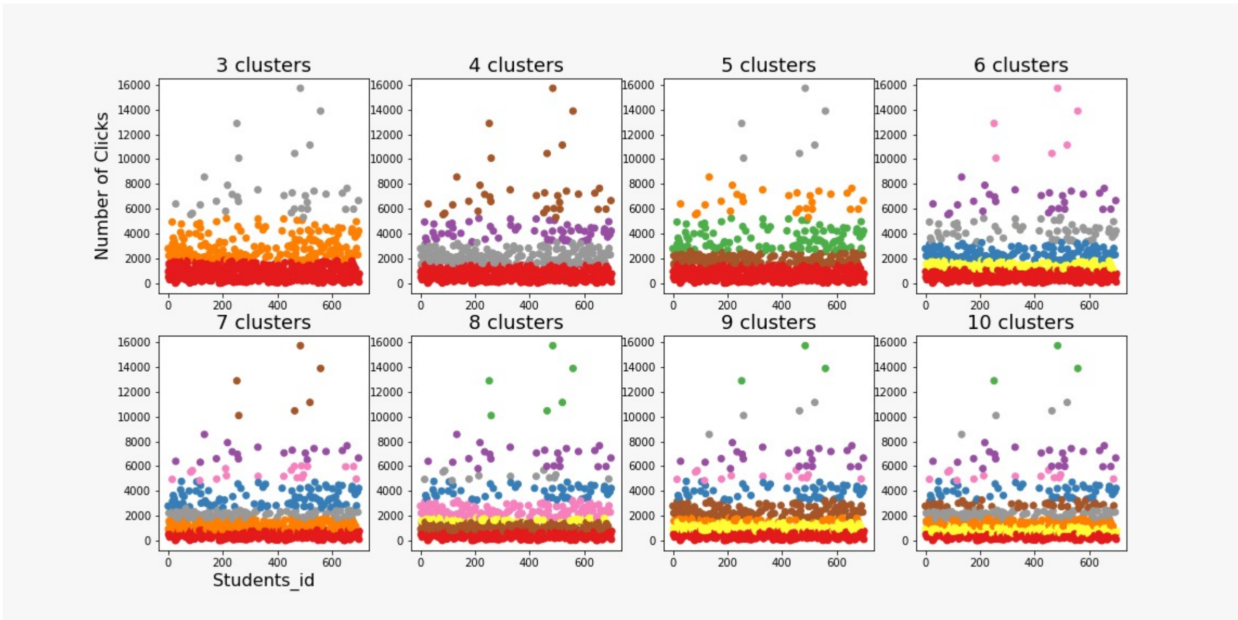
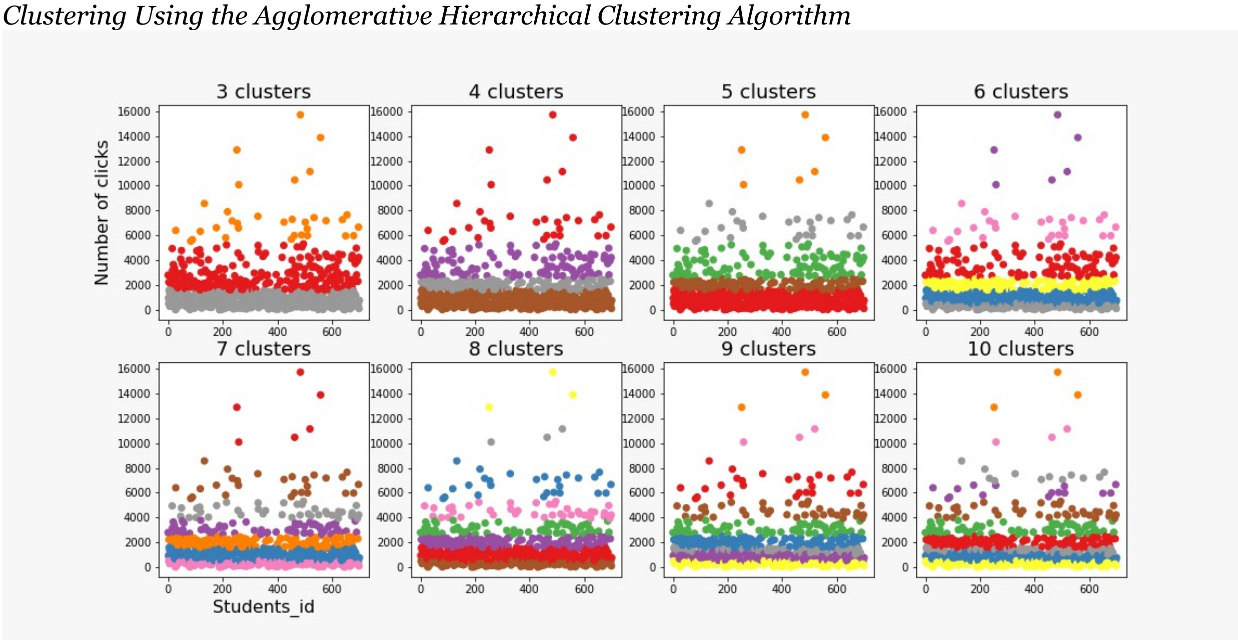


Figure 3

Clustering Using the Agglomerative Hierarchical Clustering Algorithm



Evaluation of Clustering Results

After the clusters were formed by the algorithms, an evaluation was carried out using the clValid R package that compared the cluster results and gave the optimal scores for the best performing algorithm (Brock et al., 2008). The results are presented in Table 5.

Table 5

Optimal Algorithm and Cluster Evaluation Results

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

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

The results indicate that the agglomerative hierarchical algorithm is the best performing with the optimal score of 8.4552 for Connectivity and 0.7111 for Silhouette measures when there are 3 optimal clusters. However, the Dunn index proposes 4 optimal clusters with optimal score of 0.0609. We also evaluated the clusters using the NbClust function. The NbClust function provides 30 internal validation indices that allow simultaneous evaluation of algorithms in order to determine the optimal number of clusters for a given dataset (Charrad et al., 2014). From these 30 indices, seven proposed 3 as the optimal number of clusters, fifteen proposed 4 clusters, while two proposed 5 clusters. The rest of the indices, such as the Dindex and Hubert, gave graphical results. They also indicated 4 clusters would be optimal. Based on the majority rule, we conclude that the best number of clusters in the dataset would be 4.

Self-Regulated Learning Profiles Identified from Students' Interaction Data

After the experimental evaluation of the clusters formed by agglomerative hierarchical clustering, it was revealed that the students' interaction data could optimally be categorized into four distinct clusters. The clusters seen in the dataset included:

a) Cluster 0: This cluster represented students whose number of clicks were 5,000 and over.

b) Cluster 1: This cluster represented students who had the second highest number of clicks. The range was approximately 2,500 to 5,000.

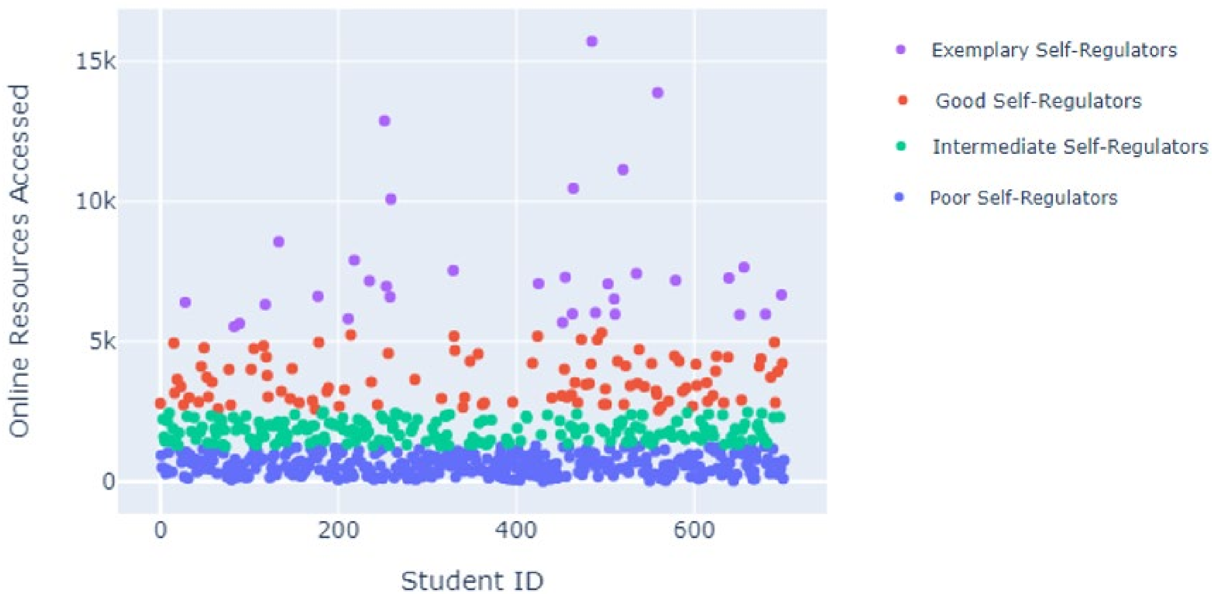
c) Cluster 2: This cluster denoted total clicks that ranged from 1,000 to 2,500.

d) Cluster 3: This cluster seemed to have similar characteristics to cluster 2 in general, and contained the lowest number of clicks, ranging from 0 to 1,000.

The classification of students into four profiles was based on behavioral activities that represented the number of resources accessed. The resources included homepage, subpages, external and internal URLs, discussion forums, course content, assignments, and course content. The SRL profiles were identified using the agglomerative hierarchical clustering algorithm. Using exploratory data analysis, the clusters formed were mapped onto four SRL profiles: exemplary self-regulators, good self-regulators, intermediate self-regulators, and poor self-regulators. These are illustrated on the scatterplot in Figure 4.

Figure 4

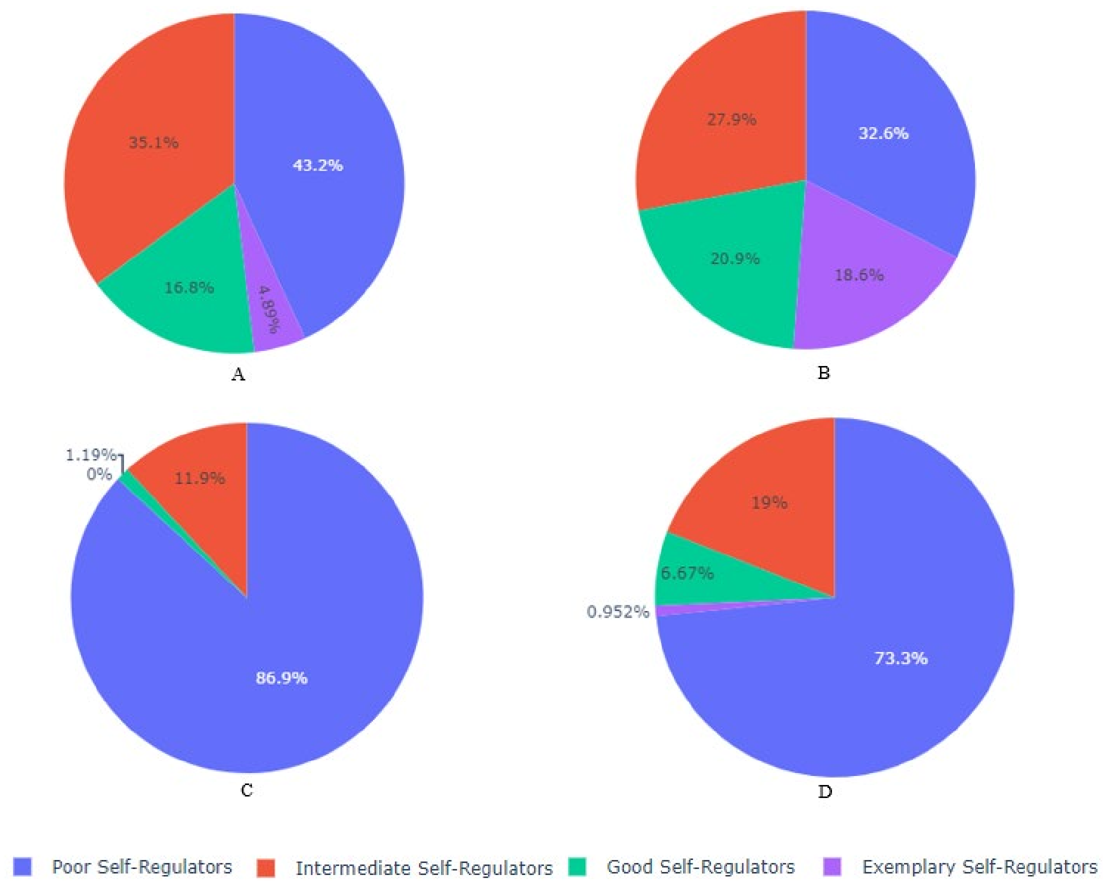
Clusters Mapped on to SRL Profiles



The original dataset included the final results of the students. Based on these results, it was possible to identify the distribution of clusters among students who had passed with distinction, passed, failed, or withdrawn. The exemplary and good self-regulators had the highest number of clickstream interactions and performed the best in terms of the final grades. The students in these two profiles either had a distinction or a pass in their final results. As presented in Figure 5A, among students who passed, 35.11% were intermediate self-regulators while 16.81% and 4.89% were good and exemplary regulators respectively. Among the students who passed with distinction, good and exemplary self-regulators represented the highest percentage at 20.93% and 32.56% respectively as illustrated in Figure 5B. The number of poor and intermediate self-regulators found among the students who had passed with distinction reveals that there could be other factors contributing to their academic performance. As shown in Figure 5B and 5C, the poor and intermediate self-regulators had a low to medium number of clickstream interactions. The majority of the students in these groups exhibited similar academic results. They either failed or withdrew from the course. It can also be observed that some students who were classified as good or exemplary self-regulators withdrew from or failed the course. This implies that there could be external factors that influenced their academic performance. Lastly, as shown in Figure 5D, among the students who withdrew, 73.33% were poor self-regulators while 19.05% represented the intermediate self-regulators.

Figure 5

Distribution of SRL Profiles Among Students Based on Their Final Results



Note. *N* of students = 735. Panel A: Students who passed. Panel B: Students who passed with distinction. Panel C: Students who failed. Panel D: Students who withdrew.

Relationship Between the SRL Profiles and the Students' Final Results

The chi-square test was carried out to establish the correlation between the SRL profiles formed by the agglomerative hierarchical clustering algorithm and students' final results. A contingency table was computed from the values of the distribution of students among the four clusters of SRL profiles and the four categories of the students' final results: passed with distinction, passed, failed, and withdrew. The computed *p*-value was 0.00 (8.988568648725134^{e-22}). When the p-value obtained is compared with the alpha value of 0.05, since $p < 0.05$, we can conclude that there is a significant relationship between the SRL profiles and the students' final results.

General Discussion

In this research, two related studies were carried out. First, a review of the literature describing EDM techniques for identifying profiles in SRL was undertaken. The results from the review indicate that a clustering technique is the most appropriate, preferred over other techniques such as temporal data mining, natural language processing, neural networks, and classification. It was observed that clustering was most often the most appropriate technique when using online educational datasets from LMSs. The findings led us to conduct the second study which aimed at experimenting with the clustering techniques such that three algorithms were compared: k-means, agglomerative hierarchical clustering, and expectation-maximization. The clustering algorithms were evaluated using internal validation measures to identify the optimal algorithms and number of clusters. The findings demonstrate that agglomerative hierarchical clustering is the best performing algorithm. These findings align with results from previous studies (Çebi and Güyer, 2020; Gašević et al., 2017). Cluster evaluation was carried out to establish the optimal algorithm with an optimal number of clusters. Using the NbClust function, where 30 inbuilt indices were used to simultaneously compare the clusters, fifteen indices proposed 4 clusters while seven indices proposed 3 clusters. Based on the majority rule, we concluded that the optimal number of clusters is four (Charrad et al., 2014). Furthermore, an exploration and analysis of the clusters formed by the optimal clustering algorithm, agglomerative hierarchical, indicate that four SRL profiles existed in the online dataset collected from a virtual learning environment. The four clusters were further examined and mapped onto four SRL profiles based on the learners' behaviors as inferred from the OULAD dataset.

The SRL profiles identified include exemplary self-regulators, good self-regulators, intermediate self-regulators, and poor self-regulators. The SRL clusters differed from each other in terms of the frequency of the sum of clicks which represents the clickstream interactions students had with online learning resources such as course homepage, external and internal URLs, course subpages, resources, discussion forums, glossary, collaboration tools, and course content. Additionally, since the OULAD dataset included students' final results, it was possible to identify the distribution of each of the profiles among the students who had distinction, pass, fail, or withdrawn. It was observed that the exemplary and good self-regulators had the highest number of clickstream interactions, i.e., above 2,500. The intermediate self-regulators had a medium number of clicks that ranged from 1,000 to 2,500, while poor self-regulators had the lowest number, i.e., below 1,000. The distribution of students in the various profiles also indicates that a majority of the poor and intermediate self-regulators either failed or withdrew from the course.

Finally, a test of independence to establish the relationship between the SRL profiles and the students' final results revealed a significant relationship between the two categorical variables. Profiling students according to their SRL skills helps instructors in identifying learners with similar interaction behaviors. These SRL profiles may be helpful in developing and providing customized and targeted interventions based of each group's characteristics.

Conclusion

Online learners differ in terms of the behaviors they exhibit during online learning. Identifying existing behavior groups will help educators provide targeted SRL interventions instead of offering one-size-fits-all treatments to students. While any algorithm can be applied to determine the number of clusters available in a given dataset, any algorithm may fail to identify the optimal number of clusters given differences in datasets. For example, datasets from educational environments differ from datasets obtained from other industries. Additionally, our review of literature revealed little knowledge exists about the most appropriate algorithm to use with datasets from online learning environments such as LMSs. This study sought to solve this problem from three perspectives: (a) the most appropriate EDM techniques being applied in identifying SRL profiles, (b) the best performing algorithm, and (c) the optimal number of SRL profiles available in trace data collected from an online learning environment.

The current study has provided insights into the identification of SRL profiles using EDM techniques such as clustering algorithms in online learning environments. The OULAD dataset was applied to the experimental comparison of the algorithms. The findings revealed that it is now possible for SRL interventions to be targeted to the right groups, based on learners' behavioral characteristics. This will enhance students' abilities in terms of SRL skills which have been found to be poor in most online learners (Goda et al., 2020). Moreover, given the large number of students enrolling in online learning and the limited number of instructors, it will be necessary to use EDM techniques to identify SRL profiles which can then be used to establish the nature and level of student interactions in online learning environments such as an LMS (Goda et al., 2020).

The findings from this study imply that EDM techniques offer great opportunities for researchers to use trace data collected from online learning environments to explore supporting SRL. Profiling learners according to their SRL strategies will be a first step in providing targeted SRL interventions. The findings from this study offer insights into two areas: first, that EDM techniques can be used to identify learner profiles in terms of SRL skills in open and distributed learning environments. Second, clustering students based on their levels of self-regulation provide a means of understanding where online learners are situated so as to develop guidance and support aligned to learners' needs hence offering the opportunity for instructors to provide targeted interventions for each of the formed clusters. The results from this study also contribute to the measuring of SRL in online learning environments by giving insights into how to build machine learning models that can ultimately be used to provide SRL interventions.

The findings concerning the association between SRL profiles and students' final results were based on correlation analysis. The results may therefore have failed to reveal all the intervening factors that could have contributed to the success or failure of the online learners. It would therefore be interesting for future studies to consider variables other than clickstream interaction behavior that could affect the clusters. Given that this current study did not consider specific SRL strategies such as time management, help-seeking, elaboration, and rehearsal, and how they could be inferred from the trace data, an empirical study could be carried out to profile learners based on identifying specific SRL strategies and examining how they could be measured, monitored, and even promoted in an actual online learning environment (Araka et al., 2021). Finally, we propose that future studies could examine how targeted interventions could be designed to promote SRL strategies based on learner needs in each SRL profile. For example, it would be interesting to investigate how EDM algorithms could be integrated into an LMS to enable real-time profiling of learners, thus providing SRL interventions to stimulate the growth of self-regulatory skills especially for poor self-regulators. Early identification and intervention will help learners with such low self-regulatory skills. We are currently carrying out an empirical study to establish whether SRL interventions provided through real-time analysis of educational data in a live LMS can improve students' learning processes and consequently advance the knowledge and behavior of learners.

Acknowledgement

This research was supported by the National Research Fund 2016/2017 grant award under the multidisciplinary-multi-institutional category to a consortium led by Kenyatta University and involving the University of Nairobi and The Cooperative University of Kenya.

References

Ainscough, L., Leung, R., Colthorpe, K., & Langfield, T. (2019). Characterizing university students' self-regulated learning behavior using dispositional learning analytics. In J. Domenech, P. Merello, E. de la Poza, D. Blazquez, & R. Peña-Ortiz (Eds.), *Fifth international conference on higher education advances (HEAd'19)* (pp. 233–241). Editorial Universitat Politècnica de València. https://doi.org/10.4995/head19.2019.9153

Aljohani, N. R., Fayoumi, A., & Hassan, S.-U. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability*, 11(24), 7238. https://doi.org/10.3390/su11247238

Alshabandar, R., Hussain, A., Keight, R., Laws, A., & Baker, T. (2018). The application of Gaussian mixture models for the identification of at-risk learners in massive open online courses. In *2018 IEEE congress on evolutionary computation (CEC 2018)* (pp. 1–8). IEEE. https://doi.org/10.1109/CEC.2018.8477770

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

Araka, E., Maina, E., Gitonga, R., Oboko, R., & Kihoro, J. (2021). University students' perception on the usefulness of learning management system features in promoting self-regulated learning in online learning. *International Journal of Education and Development Using Information and Communication Technology (IJEDICT)*, *17*(1), 45–64. http://ijedict.dec.uwi.edu/viewarticle.php?id=2850

Araka, E., Oboko, R., Maina, E., & Gitonga, R. (2020). A conceptual educational data mining model for supporting self-regulated learning in online learning environments. In J. Keengwe & Y. Tran (Eds.), *Handbook of research on equity in computer science in P-16 education* (pp. 278–292). IGI Global. https://doi.org/10.4018/978-1-7998-4739-7.ch016

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

Barnard-Brak, L., Paton, V. O., & Lan, W. Y. (2010). Profiles in self-regulated learning in the online learning environment. *International Review of Research in Open and Distributed Learning*, *11*(1), 55–78. https://doi.org/10.19173/irrodl.v11i1.769

Bosch, N., Crues, W., Henricks, G., Perry, M., Angrave, L., Shaik, N., Bhat, S., & Anderson, C. (2018). Modeling key differences in underrepresented students' interactions with an online STEM course. In A. L. Story (Chair), *Proceedings of the Technology, Mind, and Society conference 2018: TechMindSociety '18*, 1–6. https://doi.org/10.1145/3183654.3183681

Bouchet, F., Harley, J. M., Trevors, G. J., & Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 104–146. https://doi.org/10.5281/zenodo.3554613

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

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

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

Çebi, A., & Güyer, T. (2020). Students' interaction patterns in different online learning activities and their relationship with motivation, self-regulated learning strategy and learning performance. *Education and Information Technologies*, *25*, 3975–3993. https://doi.org/10.1007/s10639-020-10151-1

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

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

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

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. In A. Pardo, K. Bartimote-Aufflick, & G. Lynch (Chairs), *LAK '18: Proceedings of the eighth international conference on learning analytics and knowledge* (pp. 191–200). ACM. https://doi.org/10.1145/3170358.3170381

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

Crossley, S., Mcnamara, D., Paquette, L., Baker, R., & Dascalu, M. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In D. Gašević & G. Lynch (Chairs), *LAK '16: Proceedings of the sixth international conference on learning analytics and knowledge* (pp.

6–14). ACM. https://doi.org/10.1145/2883851.2883931

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

Di Mitri, D., Börner, D., Scheffel, M., Ternier, S., Drachsler, H., & Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. In A. Wise, P. Winne, & G. Lynch (Chairs), *LAK '17: Proceedings of the seventh international learning analytics and knowledge conference* (pp. 188–197). ACM. https://doi.org/10.1145/3027385.3027447

Di Mitri, D., Scheffel, M., Drachsler, H., Borner, D., Ternier, S., & Specht, M. (2016). Learning pulse: Using wearable biosensors and learning analytics to investigate and predict learning success in self-regulated learning. In R. Martinez-Maldonado, D. Hernandez-Leo (Eds), *Proceedings of the First International Workshop on Learning Analytics Across Physical and Digital Spaces co-located with 6th International Conference on Learning Analytics & Knowledge (LAK 2016) (pp.* 34–39). http://ceur-ws.org/Vol-1601/CrossLAK16Paper7.pdf

Elsayed, A., Caeiro-Rodríguez, M., Mikic-Fonte, F., & Llamas-Nistal, M. (2019). Research in learning analytics and educational data mining to measure self-regulated learning : A systematic review. In *Proceedings of world conference on mobile and contextual learning 2019* (pp. 46–53). https://www.learntechlib.org/p/210600/

Gašević, D., Jovanović, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, *4*(2), 113–128. https://doi.org/10.18608/jla.2017.42.10

Goda, Y., Kato, H., Yamada, M., Saito, Y., Matsuda, T., & Miyagawa, H. (2020). From adaptive learning support to fading out support for effective self-regulated online learning. In D. Glick, A. Cohen, & C. Chang (Eds.), *Early warning systems and targeted interventions for student success in online courses* (pp. 218–238). IGI Global. https://doi.org/10.4018/978-1-7998-5074-8

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

Jha, N. I., Ghergulescu, I., & Moldovan, A. N. (2019). OULAD MOOC dropout and result prediction using ensemble, deep learning and regression techniques. In H. Lane, S. Zvacek, & J. Uhomoibhi (Eds.), *Proceedings of the 11th international conference on computer supported education, Volume 2* (pp. 154–164). https://doi.org/10.5220/0007767901540164

Jo, I., Park, Y., Yoon, M., & Sung, H. (2016). Evaluation of online log variables that estimate learners' time management in a Korean online learning context. *International Review of Research in Open and Distributed Learning*, *17*(1), 195–213. https://doi.org/10.19173/irrodl.v17i1.2176

Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine, 96*(3), 118–121. https://journals.sagepub.com/doi/pdf/10.1177/014107680309600304

Khanna, L., Narayan Singh, S., & Alam, M. (2016). Educational data mining and its role in determining factors affecting students academic pPerformance: A systematic review. In *2016 1st India international conference on information processing (IICIP)* (pp. 1–7). IEEE. https://www.doi.org/10.1109/IICIP.2016.7975354

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

Kinnebrew, J., Loretz, K., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, *5*(1),190–219. https://doi.org/10.5281/zenodo.3554617

Kizilcec, R., Erez-Sanagustín, M. P., & Maldonado, J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education*, *104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *LAK '13: Proceedings of the third international conference on learning analytics and knowledge* (pp. 170–179). ACM. https://doi.org/10.1145/2460296.2460330

Klug, J., Ogrin, S., & Keller, S. (2011). A plea for self-regulated learning as a process: Modelling, measuring and intervening. *Psychological Test and Assessment Modeling*, *53*(1), 51–72.

Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., Vaclavek, J., & Wolff, A. (2015). OU analyse: Analysing at-risk students at The Open University. *Learning Analytics Review: LAK15-1*, 1–16. http://libeprints.open.ac.uk/42529/

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

Li, H., Flanagan, B., Konomi, S., & Ogata, H. (2018). Measuring behaviors and identifying indicators of self-regulation in computer-assisted language learning courses. *Research and Practice in Technology Enhanced Learning*, *13*, Article 19. https://doi.org/10.1186/s41039-018-0087-7

Lodge, J. M., & Corrin, L. (2017). What data and analytics can and do say about effective learning. *Npj Science of Learning*, *2*(1), Article 5. https://doi.org/10.1038/s41539-017-0006-5

Madni, H. A., Anwar, Z., & Shah, M. A. (2017). Data mining techniques and applications - A decade review. In J. Zhang (Ed.), *2017 23rd international conference on automation and computing*

(ICAC) (pp. 1-7). IEEE. https://doi.org/10.23919/IConAC.2017.8082090

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

Manzanares, M., Sánchez, M., García, O., & Díez-Pastor, J. (2017). How do b-learning and learning patterns influence learning outcomes? *Frontiers in Psychology*, *8*(745), 1–13. https://doi.org/10.3389/fpsyg.2017.00745

Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In S. Hsiao, J. Cunningham, K. McCarthy, G. Lynch, C. Brooks, R. Ferguson, & U. Hoppe (Chairs), *LAK '19: Proceedings of the ninth international conference on learning analytics and knowledge*(pp. 461–470). ACM. https://doi.org/10.1145/3303772.3303787

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

Nuankaew, W., Nuankaew, P., Teeraputon, D., Phanniphong, K., & Bussaman, S. (2019). Perception and attitude toward self-regulated learning of Thailand's students in educational data mining perspective. *International Journal of Emerging Technologies in Learning*, *14*(9), 34–49. https://doi.org/10.3991/IJET.V14I09.10048

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance: A case study. In D. Gašević, G. Lynch, S. Dawson, H. Drachsler, & C. Penstein Rosé (Chairs), *LAK '16: Proceedings of the sixth international conference on learning analytics and knowledge* (pp. 422–429). ACM. https://doi.org/10.1145/2883851.2883883

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*, *10*(1), 82–92. https://doi.org/10.1109/TLT.2016.2639508

Park, J., Yu, R., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). Understanding student procrastination via mixture models. In K. E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th international conference on educational data mining* (pp. 187–197). International Educational Data Mining Society.

Peach, R. L., Yaliraki, S. N., Lefevre, D., & Barahona, M. (2019). Data-driven unsupervised clustering of online learner behaviour. *Npj Science of Learning*, *4*(1), Article 14. https://doi.org/10.1038/s41539-019-0054-0

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Rodriguez, A., Vázquez Barreiros, B., Lama, M., Gewerc, A., & Mucientes, M. (2014). Using a learning analytics tool for evaluation in self-regulated learning. *Proceedings of the 2014 Frontiers in Education conference* (pp. 2484–2491). IEEE. https://doi.org/10.1109/FIE.2014.7044400

Rodriguez, F., Rivas, M. J., Yu, R., Warschauer, M., Park, J., & Sato, B. K. (2019). Utilizing learning analytics to map students' self-reported study strategies to click behaviors in STEM courses. In S. Hsiao, J. Cunningham, K. McCarthy, G. Lynch, C. Brooks, R. Ferguson, & U. Hoppe (Chairs), *LAK '19: Proceedings of the ninth international conference on learning analytics and knowledge* (pp. 456–460). ACM. https://doi.org/10.1145/3303772.3303841

Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLoS ONE*, *14*(1), Article e0210236. https://doi.org/10.1371/journal.pone.0210236

Saadati, Z., Zeki, C. P., & Barenji, R. V. (2021, April 29). On the development of blockchain-based learning management system as a metacognitive tool to support self-regulation learning in online higher education. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1920429

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

Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In S. B. Shum, D. Gašević, & R. Ferguson (Chairs), *LAK '12: Proceedings of the second international conference on learning analytics and knowledge* (pp. 252–254). ACM. https://doi.org/10.1145/2330601.2330661

Silvola, A., Näykki, P., Kaveri, A., & Muukkonen, H. (2021). Expectations for supporting student engagement with learning analytics: An academic path perspective. *Computers & Education*, *168*, Article 104192. https://doi.org/10.1016/j.compedu.2021.104192

Sun, Z., Lu, L., & Xie, K. (2016). The effects of self-regulated learning on students' performance trajectory in the flipped math classroom. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The international conference of the learning sciences (ICLS) 2016, Volume 1* (pp. 66–73). International Society of the Learning Sciences. https://www.isls.org/icls/2016/docs/ICLS2016_Volume_1_30June2016.pdf

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

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

Valdiviezo, P., Reátegui, R., & Sarango, M. (2013). Student behavior patterns in a virtual learning environment. In M. M. Larrondo Petrie, H. Alvarez, I. E. Esparragoza, & C. Rodriguez Arroyave (Eds.), *Innovation in engineering, technology and education for competitiveness and prosperity: Proceedings of the 11th Latin American and Caribbean conference for engineering and technology* (pp. 1–8). LACCEI. http://www.laccei.org/LACCEI2013-Cancun/RefereedPapers/RP091.pdf

Valle, A., Núñez, C., Cabanach, R., González-Pienda, J., Rodríguez, S., Rosário, P., Cerezo, R., & Muñoz-Cadavid, M. (2008). Self-regulated profiles and academic achievement. *Psicothema*, *20*(4), 724–731. http://www.psicothema.com/pdf/3547.pdf

Van-Craenendonck, T., & Blockeel, H. (2015, June 19). *Using internal validity measures to compare clustering algorithms* [Poster presentation]. Benelearn, Delft, Netherlands. https://lirias.kuleuven.be/1656512?limo=0

Winne, P., & Baker, R. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 1–8. https://doi.org/10.5281/zenodo.3554619

Wong, J., Baars, M., de Koning, B. B., & Paas, F. (2021). Examining the use of prompts to facilitate self-regulated learning in massive open online courses. *Computers in Human Behavior*, *115*, Article 106596. https://doi.org/10.1016/j.chb.2020.106596

Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers and Education*, *140*, Article 103595. https://doi.org/10.1016/j.compedu.2019.103595

Yot-Domínguez, C., & Marcelo, C. (2017). University students' self-regulated learning using digital technologies. *International Journal of Educational Technology in Higher Education*, *14,* Article 38. https://doi.org/10.1186/s41239-017-0076-8

Yu, R., Jiang, D., & Warschauer, M. (2018). Representing and predicting student navigational pathways in online college courses. In R. Luckin, S. Klemmer, & K. Koedinger (Chairs), *L@S '18: Proceedings of the fifth annual ACM conference on learning at scale* (pp. 1–4). ACM. https://doi.org/10.1145/3231644.3231702

Zheng, J., Xing, W., Zhu, G., Chen, G., Zhao, H., & Xie, C. (2020). Profiling self-regulation behaviors in STEM learning of engineering design. *Computers & Education*, *143*, Article 103669. https://doi.org/10.1016/j.compedu.2019.103669

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2