Zhao, F., Liu, G. Z., Zhou, J., & Yin, C. (2023). A Learning Analytics Framework Based on Human-centered Artificial Intelligence for Identifying the Optimal Learning Strategy to Intervene Learning Behavior. *Educational Technology & Society*, *26*(1), 132–146. https://doi.org/10.30191/ETS.202301_26(1).0010

A Learning Analytics Framework Based on Human-Centered Artificial Intelligence for Identifying the Optimal Learning Strategy to Intervene in Learning Behavior

Fuzheng Zhao[1,4], Gi-Zen Liu[2], Juan Zhou[3] and Chengjiu Yin[1*]
[1]Kobe University, Japan // [2]National Cheng Kung University, Taiwan // [3]Tokyo Institute of Technology, Japan // [4]Jilin University, China // zhaofz635@gmail.com // gizen@mail.ncku.edu.tw // juan.z.kt@gmail.com // yin@lion.kobe-u.ac.jp
[*]Corresponding author

ABSTRACT: Big data in education promotes access to the analysis of learning behavior, yielding many valuable analysis results. However, with obscure and insufficient guidelines commonly followed when applying the analysis results, it is difficult to translate information knowledge into actionable strategies for educational practices. This study aimed to solve this problem by utilizing the learning analytics (LA) framework. We proposed a learning analytics framework based on human-centered Artificial Intelligence (AI) and emphasized its analysis result application step, highlighting the function of this step to transform the analysis results into the most suitable application strategy. To this end, we first integrated evidence-driven education for precise AI analytics and application, which is one of the core ideas of human-centered AI (HAI), into the framework design for its analysis result application step. In addition, a cognitive load test was included in the design. Second, to verify the effectiveness of the proposed framework and application strategy, two independent experiments were carried out, while machine learning and statistical data analysis tools were used to analyze the emerging data. Finally, the results of the first experiment revealed a learning strategy that best matched the analysis results through the application step in the framework. Further, we conclude that students who applied the learning strategy achieved better learning results in the second experiment. Specifically, the second experimental results also show that there was no burden on cognitive load for the students who applied the learning strategy, in comparison with those who did not.

Keywords: Learning analytics framework, Analysis result application, Human-center AI, Learning strategy

1. Introduction

Learning analytics (LA) is viewed as a domain that combines data analytics and human judgment (Siemens, 2013). LA aims to reveal hidden patterns and generate actionable intelligence, which could provide timely intervention for students' learning behavior. A small number of efforts in LA have focused on the predictive analytics realm, in which techniques such as machine learning and deep learning were drawn upon (Xing & Du, 2018), and some analysis results have been applied to education for intervening learning behavior (Zhao et al., 2021). Some studies have shown that early prediction could help improve learning engagement (Gray & Perkins, 2019). In addition, systematic intervention strategies can successfully reduce the dropout rate (Choi et al., 2018).

While the predictive analytics domain has seen a surge, little is known about how to identify and apply the most suitable intervention strategy to students' learning behavior. Some concerns regarding intervention strategies after prediction analyses have been raised, as the application of prediction analysis results failed to show the expected efficacy. Bowers and Sprott (2012) found that some indicators may accurately predict which factors most affect academic performance but are unable to support effective interventions, owing to the lack of appropriate and effective interventions. Moreover, faced with the same analysis results of learning behavior patterns, the impact on students' learning behavior largely varies and is dependent on various intervention strategies (Rienties et al., 2016). As a result, a suggestion has been made that predictive analysis should go beyond simply predicting learning performance and should also inform instructors on appropriate intervention strategies (Barry & Reschly, 2012).

The application of analysis results is regarded as the final stage of the LA framework and is responsible for remedial actions for students (Clow, 2012). In previous research, it was found that the application step in the current LA frameworks (Campbell et al., 2007; Chatti et al., 2012; Dron & Anderson, 2009; Elias, 2011; Siemens, 2013), including the ReCoLBA framework that we proposed (Zhao et al., 2021), lacks guidance on how to transform the analysis results into corresponding implementable strategies, especially learning strategies.

Moreover, human-centered Artificial Intelligence (HAI) has been emerging as a new development trend. It not only seeks to consider the human condition when designing the AI but also identifies human learning patterns and facilitates timely intervention by artificial intelligence (AI) techniques and big data (Tsai et al., 2020). In particular, it requires the computation and application of AI algorithms with machine intelligence to be trustworthy and responsible, thereby enhancing the welfare of humankind. To increase the trustworthiness of analysis results produced by AI algorithms, make them responsible for computation, and consider human welfare when applying them, we proposed an LA framework that considers HAI, emphasizing the result application step.

2. Literature review

2.1. Opportunities of HAI for the LA framework

The AI research goals in relation to education consist of prediction, structure discovery, and relationship mining (Yin & Hwang, 2018). Considering the new development trend of AI and its typical application, machine learning, Shneiderman (2020) envisions the use of HAI as inevitably on the rise and taking an evolutionary direction, considering human conditions and contexts in its design and application. As an important branch of HAI, the essence of LA is to apply big data and AI to identify at-risk students and intervene promptly. In particular, HAI requires explainable, trustworthy, and responsible computation for AI algorithms and applications with machine intelligence, thereby improving human welfare (Yang et al., 2021).

2.2. Problems in LA frameworks

Although studies of LA frameworks have been undertaken, they have not examined the details of the implementation of the analysis result application in those frameworks. Campbell et al. (2007) initially proposed the "Five Steps of Analytics" framework, comprising the steps "capture," "report," "predict," "act," and "refine." The term "act" refers to the application of analysis results. As a type of intervention, they suggest supporting the above applications with a personal phone call or e-mail.

As the introduction of LA to various learning environments increased, different perspective-based explorations on the framework also saw urgent development. Dron and Anderson (2009) defined their "Collective Application Model" framework taking into consideration the characteristics of e-learning. As it highlighted information gathering, processing, and presentation, the analysis result application step was excluded from the framework, which comprised only "select," "capture," "aggregate," "process," and "display." After a comparison of existing frameworks, Elias (2011) proposed a comprehensive framework comprising "select," "capture," "aggregate and report," "predict," "use," "refine," and "share." Regarding the application step, it only provides a brief description of attempts to improve the learning system.

The Chatti framework (Chatti et al., 2012) incorporates analysis and its corresponding application into one step to compose the processing step, with the remaining two steps pre-processing and post-processing. In contrast, Siemens (2013) not only added an application step to his framework but also highlighted the purposes of the step, such as "intervention," "optimization," "alters," "warning," "guiding" / "nudging," "systemic," and "improvement." Two issues exist in these frameworks. First, the reason for the analysis results cannot be sufficiently understood due to the lack of confirmation of the analysis results by AI algorithms, reducing the explanation and trustworthiness of analysis results. Second, existing frameworks cannot provide specific ways to apply the analysis results of AI algorithms, as shown in Table 1. In this context, analysis results cannot be successfully translated into the proven application strategy and thus cannot effectively support human welfare. The former was addressed in the ReCoLBA framework (Zhao et al., 2021), and the latter is the focus of this study.

Table 1. Features and drawbacks of the application step in the existing frameworks

Title	Features	Drawbacks
"Five Steps of Analytics" framework	Specific intervention approaches (phone calls or e-mail)	Limited application range
"Collective Application Model" framework	Skips the application step	Lack of function
Elias' framework	Only defines the application	No specific guidance steps
Chatti's framework	Considers the application a sub-step of data processing; no other details are available	The independent properties of the application are not established
Siemens' framework	Describes the purposes of the application	Lack of application methods
ReCoLBA framework	Describes the stakeholders of the application	Lack of specific guidance

2.3. Low application effect in analysis result application

The application of analysis results in LA is a key step, aiming to apply the analysis results to educational practice for monitoring, prediction, intervention, assessment, adaptation, personalization, recommendation, and reflection (Chatti et al., 2012). To achieve the intended application targets, it is critical to fit the analysis results into implementations of education activities when developing an application.

However, having good analysis results is not always successful in facilitating educational activities, and it is especially crucial to apply them correctly (Viberg et al., 2018). Hanna (2004) attributes this to the ambiguity of the analysis results: the results may yield useful insights and clues without providing definitive answers. Furthermore, Saks et al. (2018) found that the ambiguity of the analysis results has caused most users to hold an ambiguous view of application effectiveness. On the other hand, most studies have concentrated on how to meet application requirements along with the advent of these learning scenarios, tools, and data. In contrast, little attention has been given to concrete steps to use the analysis results to create an optimal application strategy.

2.4. Big data-driven education application strategy development

The term evidence-driven education (EDE) was first used by Hargreaves (1997) and was inspired by evidence-driven practice in the field of medicine. EDE aims to bridge the research-to-practice gap in teaching as well as to shift the driving force of instructional programs and practices to evidence rather than ideology, faddism, politics, and marketing (Davies, 1999). EDE is supported by the findings of multiple, high-quality, experimental studies (Cook et al., 2008) and quantitative analysis (Moran & Malott, 2004), which provides sound evidence that an educational practice truly works. Kuromiya et al. (2020) used evidence obtained from reading behavior analysis conducted when students were learning via a learning management system to evaluate whether a new learning intervention has a positive learning effect.

In summary, significant scope for exploration remains in the application of analysis results, especially in the context of current LA frameworks. To make up for the drawbacks of the application step in all frameworks, an LA framework based on HAI, also called the HAILA framework, was designed to identify the optimal learning strategy and provide an application-step sequence.

The research questions (RQs) to be addressed in this study are as follows:

- RQ1. Is the proposed framework conducive to effectively identifying the optimal learning strategy? If so, how?
- RQ2. What is the effect of the learning strategy identified by the proposed framework?

3. Research on the HAILA framework

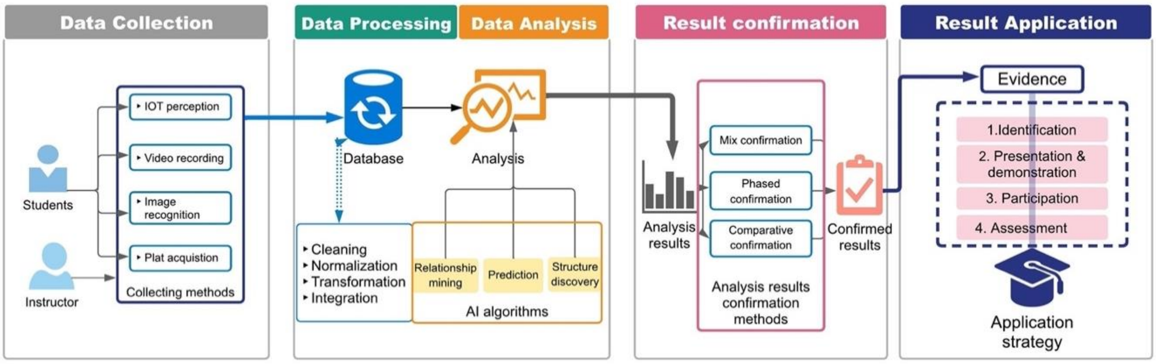
The HAILA framework is based on but differs from the ReCoLBA framework, as shown in Table 2. This framework also contains steps related to data collection, data processing, data analysis, result confirmation, and result application, as shown in Figure 1. More importantly, the HAI concept is introduced for the design process of this framework, with a focus on refining the application step to identify the optimal learning strategy to intervene in learning behavior. This framework aims to increase the dependability of AI algorithms' analysis and results, ensure the accountability of the application of their analysis results, intervene in educational practice in a timely and accurate manner, and improve learning outcomes.

Table 2. Differences and similarities between ReCoLBA and HAILA frameworks

Name	Differences	
ReCoLBA	Purpose	Confirming analysis results produced by AI algorithms
	Design concept	Does not consider HAI-driven design
	Disadvantage	Lacking specific guidance in the application-step sequence
HAILA	Purpose	Confirming analysis results by AI algorithms and finding the application strategy
	Design concept	Considers HAI-driven design
	Advantage	Has a specific application-step sequence

Diverging from other frameworks that do not consider HAI, the HAILA framework not only includes a result confirmation step to identify the accuracy of analysis results produced by AI algorithms but also provides an application-step sequence to transform the complex, abstract results obtained from AI algorithms into an application strategy. Thus, instructors' understanding of AI algorithms' analysis results will increase (Zhao et al., 2021), and the effectiveness of AI algorithm analysis applications is expected to be guaranteed.

Figure 1. HAILA framework



3.1. Data collection

In the LA field, data collection refers to a process in which information is gathered from various educational environments using a variety of techniques, such as video recording, image recognition, platform acquisition, and IOT perception. To provide a sound data foundation for successive steps, data collection should be characterized by timeliness, consistency, accessibility and convenience, accuracy, and responsiveness (Russell & Taylor, 2008).

3.2. Data processing

This consists of imputation of missing values, data noise identification, data integration, data cleaning, data normalization, and data transformation. This step aims to offer the most suitable data for analysis through the steps of retrieving, identifying, manipulating, modifying, and replacing.

3.3. Data analysis

The analysis goals include prediction, structure discovery, and relationship mining. To accomplish these goals and consider the characteristics of various education scenarios, a total of 12 methods under three categories corresponding to the above goals are applied in practice. Four methods, namely association rule mining, correlation mining, sequential pattern mining, and causal data mining, support the relationship mining goal. The structure discovery goal can be achieved using clustering, factor analysis, knowledge inference, and network analysis. The prediction goal primarily depends on classification, regression, latent knowledge, and estimation.

3.4. Result confirmation

Confirmation of the analysis results is performed before application. The reasons for the analysis results can be determined based on the confirmation methods included in the framework design. The confirmation step comprises mixed, phased, and comparative confirmation.

3.5. Result application

As part of the curriculum design, learning strategies are primarily responsible for the realization and completion of the instruction objectives. To find the optimal application strategy, an evidence-driven education policy was employed in the application-step sequence, which is responsible for building a corresponding link between the identified learning strategy and the analysis result. The evidence-driven education policy establishes the hypothesis of matching between the analysis results and the application strategy. Hence, the application-step sequence helps instructors determine the optimal application strategy. In addition, the evidence for constructing hypotheses using the optimal applied strategy is based on the analysis of data from any learning scenario and is not for one specific such scenario. Therefore, the application strategy obtained can be applied to any learning scenario. Four steps were designed as part of the application-step sequence, including identification, presentation and demonstration, participation, and assessment. The hypothesis regarding the analysis results and application strategies is fulfilled in the identification step, which reflects the design concept of evidence-based education.

3.5.1. Identification

As the existing frameworks lack specific application guidance, they do not provide the function of strategy identification. This step aims to determine an optimal learning strategy suitable for the confirmed analysis results. Thus, the degree to which the identified learning strategy matches the confirmed analysis results determines the application effect.

3.5.2. Presentation and demonstration

This step allows the instructor to disseminate information to learners using verbal information in writing, and visual symbols. It aims to gain learners' attention, inform learners of objectives, and combine new strategies with prior knowledge. When illustrating, concise and concrete steps for implementing a strategy are crucial.

3.5.3. Participation

This application step allows the learner to use the identified strategy to affect learning achievement. In this step, learners are asked to participate in a learning scenario and retry the strategy until they can use it successfully in real-world practice (Dick et al., 2015).

3.5.4. Assessment

This step has two dimensions. The first is to provide accurate feedback regarding learner performance when using this strategy, focusing on academic achievement in regard to the learning content (Dick et al., 2015). The second is to measure the cognitive load on the learner, which can reveal the impact of the learning strategy. The combination of academic performance and cognitive load measures is considered to provide a reliable estimate of the efficiency of instruction methods (Paas et al., 2003).

4. Experiment design

To examine the effectiveness and contribution of the HAILA framework, a case study was conducted using two independent experiments. The first experiment is designed to analyze reading behavior data and to confirm the trustworthiness of its analysis results. The learning behavior that contributes most to learning achievement in the experiment is expected to be determined and will be used as evidence for an application strategy to intervene in learning behavior. To verify the effectiveness of SQ3R, which is hypothesized based on the evidence that SQ3R could encourage students to turn back to re-read pages, another experiment is conducted.

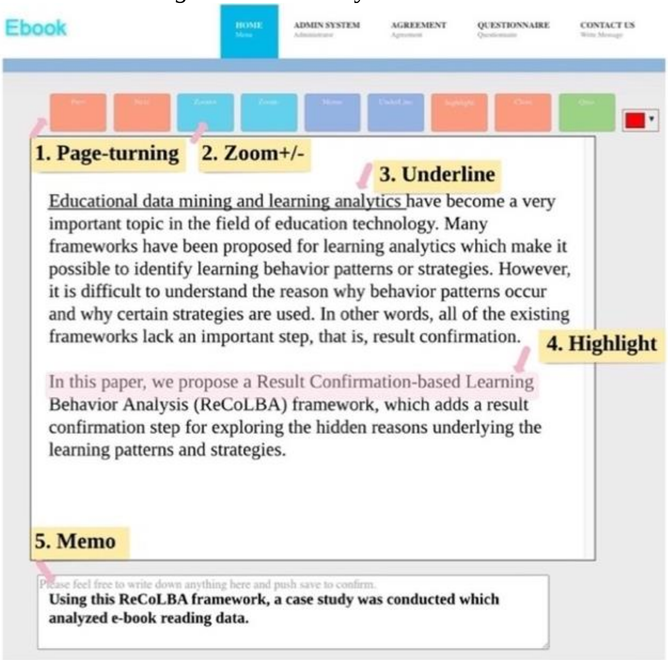
4.1. Experiment for confirming framework effectiveness

A total of 234 freshmen were recruited, among which the gender distribution was 65 males and 169 females, with an average age of 19 years old. Five experimental steps were employed according to the HAILA framework, namely (1) collecting students' reading behaviors using an e-book system, (2) using a machine learning library (scikit-learn) to process raw data, (3) utilizing classification prediction algorithms and feature importance calculation methods in the analysis step, (4) adopting different algorithms to confirm the analysis results, and (5) applying an identified learning strategy.

4.1.1. Data collection by e-book system

An e-book system was developed (Yin et al., 2018) to capture learners' reading learning behavior. Learners can conveniently read materials using several operating tools, such as (1) turning the page forward or backward, (2) resizing the view by zoom, (3) creating a memo, (4) adding or removing underlining in the learning material, and (5) adding or deleting highlights with a variety of color options, as shown in Figure 2.

Figure 2. E-book system interface



All the learning behavior observed in the learning activities was recorded in a data log with 12 data features, as shown in Table 3. Among the e-book features, the most used were the tools associated with basic reading behaviors, such as Next (NE), Prev (PR), Highlight (HL), Underline (UL), Memo (ME), Bookmaker (BM), Read time (RT), and Read page (RP). In addition, the BacktrackRate (BR) feature is a statistical ratio dividing Next and Prev, which provides a new view on repeated learning behavior. The equipment used for learning, such as a PC, mobile phone (Mobile), and tablet, was adopted as a parameter.

Table 3. Samples of reading behavior data derived from the e-book system

Id	PC	Mobile	Tablet	BR	ME	HL	UL	PR	NE	RT	RP	BR
1	0	1	0	7	1	29	3	9	12	60	51051	0.75
2	0	0	1	8	0	20	4	79	96	107	42043	0.823

4.1.2 Data processing using scikit-learn

The data processing steps were divided into two categories. The first consisted of basic processing, mainly involving the removal of missing values to maximize the validity of the data and standardization to unify the values of each data feature. As a result, valid data from 229 participants remained, and all data values were compressed from 0 to 1. The second category focused on the analysis method–specific data preparation. Considering the demands of data balance for binary classification prediction (Krawczyk, 2016), we adjusted the passing score of 60, which had a huge gap in proportion, to 70. In addition, splitting data into train- and test-datasets is necessary as part of regular data processing. The *train_test_split* function built into scikit-learn was adopted to achieve the specific splitting ratio of 3:7, in which the training dataset accounted for 70%.

4.1.3. Data analysis by decision tree model

In this study, the decision tree model was used to explore which data feature contributes most to predicting academic performance (Hamoud et al., 2018; Mesarić & Šebalj, 2016). After the prediction model was created and tuned by scikit-learn, we obtained the final model with five metrics: accuracy (0.739), F1-score (0.763), recall (0.763), precision (0.763), and AUC (0.81). Following the rule of thumb, the predictive performance of this model can be viewed as fair. Note that an AUC score over 0.8 represents good comprehensive performance in terms of a model's effectiveness.

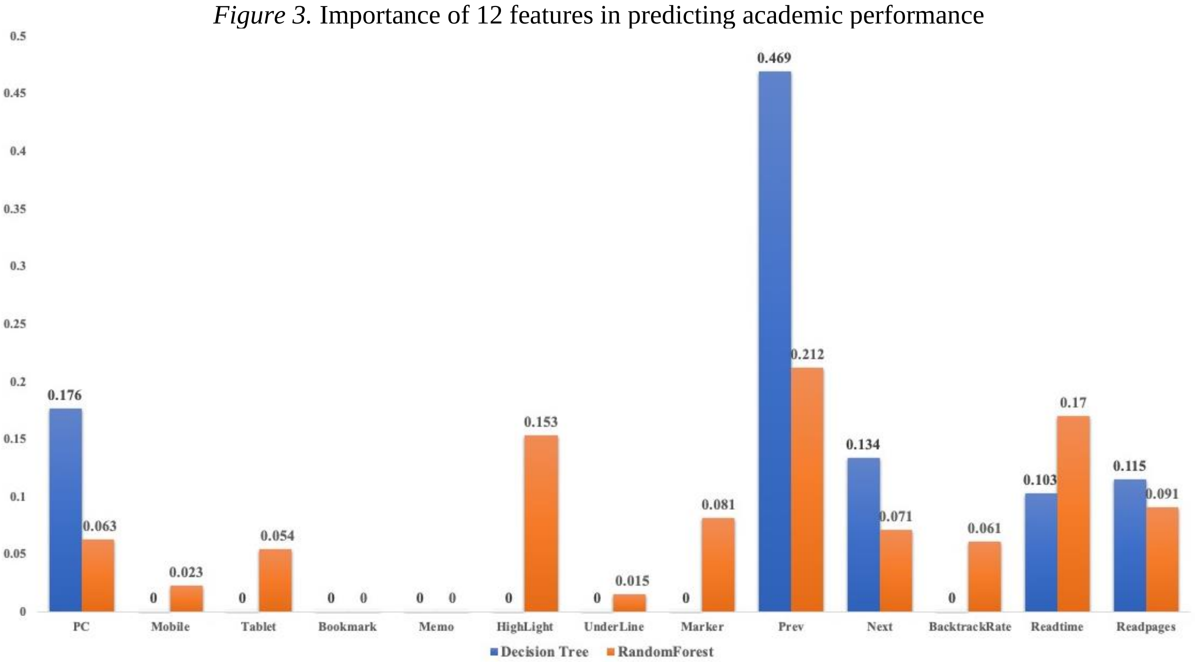
Scikit-learn offers an impurity-based feature importance calculation function oriented to the decision tree model. Subsequently, the 12 data features were analyzed using this function. As fundamental splitting parameters to calculate the feature importance for classification prediction, the Gini Index and Information Gain prevail in terms of splitting criteria. Raileanu and Stoffel (2004) found that there are no obvious differences by comparing the efficiency of splitting features for tree models. Moreover, the Gini Index has proved to be better than the other splitting parameters specifically for unbalanced datasets (Park & Kwon, 2011). As our sample size of students who fail the exam is unbalanced relative to those who passed, accounting for a minority, the Gini Index was adopted to calculate the probability weighting of each node in the tree model, with values ranging between 0 and 1, where 0 expresses the purity of classification. The values for PC, Prev, Nex, Read time, and BacktrackRate are 0.176, 0.496, 0.134, 0.103, and 0.115, respectively, and the other features have values of 0. It was found that the Prev feature had the greatest impact on prediction performance. In other words, students who have Prev learning behavior are more likely to pass the exam and obtain better academic performance.

4.1.4. Result confirmation by a comparative method

The most effective data feature, Prev, was successfully identified for predicting students' academic performance. However, it was unclear whether this analysis result was sufficiently accurate to obtain the same results in other algorithms. Guided by the analysis result confirmation step, a comparative confirmation method was employed to confirm the correctness and reproducibility of the analysis results. As an innovative algorithm based on the tree model (Liaw & Wiener, 2002), the random forest algorithm outperforms the other algorithms (Breiman, 2001) and is commonly used for binary prediction models.

In the confirmation step, the same experimental conditions were set, and only the selected algorithm was shifted from the decision tree to the random forest model. Following the rule of thumb, the prediction based on the random forest model was also acceptable in terms of accuracy (0.753), precision (0.729), recall (0.794), F1-score (0.76), and AUC (0.75) according to the results of data feature importance in the decision tree and random forest models, as shown in Figure 3. This shows that the Prev parameter consistently contributes the most in the prediction of academic performance, indicating that students who show the Prev learning behavior have a higher probability of passing the exam.

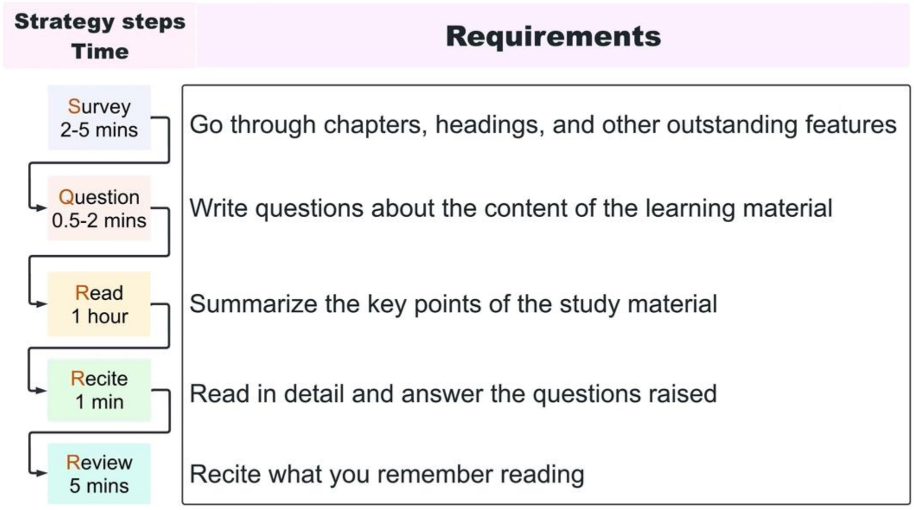
Figure 3. Importance of 12 features in predicting academic performance



4.1.5. Result application by SQ3R

After the researchers analyzed learning behavior using AI algorithms, the learning behavior that contributed most to learning efficiency is to be considered evidence of strategy making. Moreover, a hypothesis regarding application strategy was presented to facilitate the improvement of learning efficiency. First, the identification step was entered. The Prev behavior raised the need to increase learning frequency. Consistent with the demand in the learning frequency, we found that the SQ3R learning strategy was primarily designed for university students reading academic textbooks (Li et al., 2014). It is a comprehensive reading method comprising five steps: survey, question, read, recite, and review (Flippo & Bean, 2018). Importantly, multiple SQ3R learning steps can increase the occurrence of repeated learning behaviors (Huber, 2004). As shown in Figure 4, short-term memory achieved by repeatedly turning pages is necessary when students are quickly browsing notable features and writing down reading questions in survey and question steps. In subsequent steps, reciting and confirming what has been read are also realized by turning pages. To this end, it is hypothesized that SQ3R facilitates the occurrence of the Prev behavior.

Figure 4. SQ3R schedule



In the presentation and demonstration step, an introduction of SQ3R explaining how to use it in the e-book system was provided by an instructor. In Figure 4, a schedule was also designed based on previous instruction experience. For example, the survey step is recommended to last 2 to 5 min, including reviewing notable features in the textbook by turning pages and highlighting or underlining keywords, followed by 30 seconds to 2 mins to write down questions as memos, 1 hour for reading in detail, 1 min for reciting the memory, and 5 mins for summarizing and reviewing the material that was the emphasis of the learning experience by quickly turning pages. Subsequently, learners could undertake the participation step with the guidance of the presentation and demonstration steps. Finally, a post-test regarding the learning content and a test of cognitive load provide feedback to learners in the assessment step.

4.2. Experiment for verifying the framework contribution

This study examines the contribution of this framework by exploring the effectiveness of the SQ3R. Particularly, this experiment aims to evaluate the effectiveness of the SQ3R in helping students improve their academic performance. To this end, an e-book system was used to complete the above experiment to evaluate whether significant changes in academic performance and reading behaviors were related to the SQ3R application.

4.2.1. Participants

Thirty-seven male and 32 female freshmen, aged from 19 to 21 years, participated in this study. All participants were randomly assigned to two groups: the experimental group and the control group. The 35 participants in the experimental group were asked to read a learning material using the SQ3R, while the control group, which comprised 34 participants, was assigned to read the learning material without the guidance of the SQ3R.

4.2.2. Measuring method

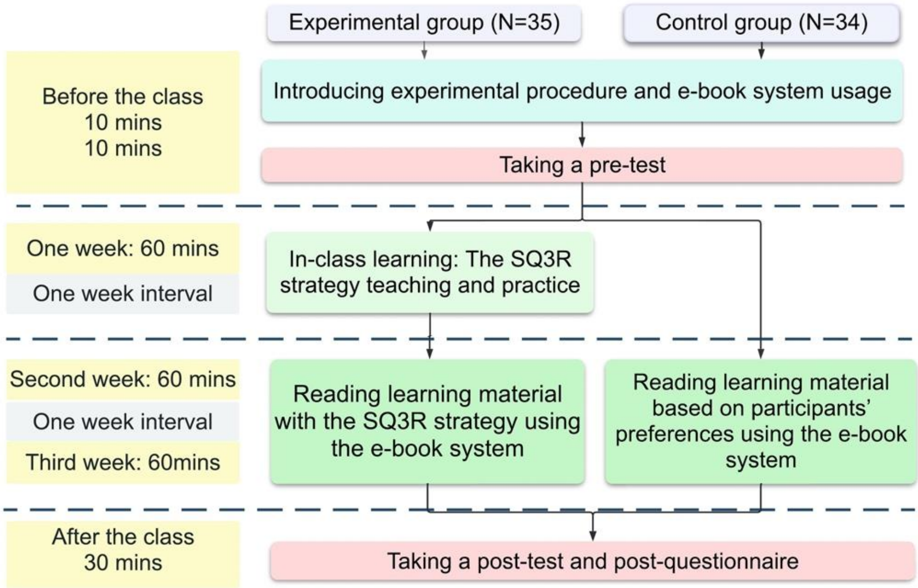
The measuring method comprised a pretest and a post-test. Before the experiment, a pretest was conducted to assess whether the two groups had the same equivalent prior knowledge regarding the learning content and the SQ3R. This pretest consisted of 10 multiple-choice items. The post-test aimed to measure whether the SQ3R was beneficial to participants' academic performance. This test was similar to the pre-test, comprising 10 multiple-choice items related to the learning emphasis in the upcoming learning materials. The pretest and post-test were both scored on a 10-point scale.

A post-questionnaire was employed to identify participants' cognitive load when using the SQ3R, and their attitude about introducing new learning technologies into the reading environment. Based on the measurement created by Sweller (1988), a questionnaire to investigate cognitive load was modified. The developed questionnaire consisted of eight questions with two dimensions: mental load and mental effects. Each question was scored on a 5-point scale, where 5 represented "strongly agree" and 1 indicated "strongly disagree." The Cronbach's alpha values of the two dimensions were 0.91 and 0.95.

4.2.3. Experimental procedure

According to Figure 5, this study consisted of a pretest, a reading-based learning activity using the e-book system, a post-test, and a post-questionnaire on cognitive load and learning strategy acceptance, which all together took 2 weeks.

Figure 5. Experimental procedure



Initially, an orientation was provided for participants to introduce the experimental procedure and the e-book system operation and precautions. Following the orientation, the experimental and control groups were asked to complete a pretest to evaluate their knowledge of the SQ3R. Afterward, in-class instruction of the SQ3R was provided only to the experimental group over 60 mins during the first class, followed by an independent practice targeting the SQ3R proficiency during 1 week. Then, the experiment proceeded to two learning activities that took 60 mins each, with an interval of 1 week in between. The participants in the experimental group read the learning material using the SQ3R, whereas those in the control group read the same learning material, but the reading strategy was based on their preferences. Subsequently, a post-test and post-questionnaire, which took 30 mins, were conducted with the two groups.

5. Experimental results

We first explored the learning behavior that contributes most to learning achievement by analyzing learning behavior in an e-book system and used the analysis results as evidence for developing application strategies, hypothesizing that the developed strategies could promote learning achievement. Finally, the hypothesis has been verified by the following experiment. Hypothesis testing utilizing 61 valid samples, including an independent sample *t*-test and one-way analysis of variance (ANOVA) in SPSS Statistics, was employed to examine the effectiveness of the SQ3R in improving academic performance and affecting learning behavior.

The experimental results show that the experimental group using the SQ3R significantly outperformed the control group in terms of academic performance, even with equivalent prior knowledge. Furthermore, regarding

Prev, Read time, and Read page behaviors, the experimental group showed a significantly higher number of occurrences. In addition, the above differences are also verified by data visualization. Finally, the cognitive load test revealed that the SQ3R did not impose additional cognitive load on students' learning.

5.1. Analysis of academic performance

The pretest results showed that the standard deviation and mean values were 2.18 and 6.67 for the experimental group and 2.15 and 7.35 for the control group. According to the *t*-test results (t = -1.23, p > .05; Table 4), no statistically significant differences were found between the two independent groups. Thus, all participants in both groups were known to have equivalent prior knowledge regarding the SQ3R.

After the learning activity, ANOVA was conducted to determine whether there was any statistically significant difference in the post-test results between the two groups. Using the groups as a fixed factor and post-test scores as the dependent variable, a common assessment for homogeneity of variance was performed using Levene's test. A Levene's test score above 0.05 (F = .09, p = .75) indicated that this test is robust to violations of the assumption. It was concluded that the experimental group was statistically different from the other group. Based on the statistical results (F = 32.86, p < .01) shown in Table 5, the participants who learned with the SQ3R showed significantly better academic performance than those who did not. In other words, the SQ3R was helpful for participants in improving their academic performance.

Table 4. Descriptive data and *t*-test result of the pretest results

Variable	Group	*N*	Mean	*SD*	*t*
Pretest	Control group	30	6.67	2.18	-1.23^{*}
	Experiment group	31	7.35	2.15	

Note. $^{*}p$ > .05.

Table 5. Descriptive data and ANOVA result of the post-test results

Variable	Group	*N*	Mean	*SD*	*F*
Post-test	Control group	30	4.50	1.77	32.86^{**}
	Experiment group	31	7.06	1.71	

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

5.2. Analysis of learning behavior

First, a *t*-test was used to analyze the repeated learning behavior based on Prev, Read time, and Read page in the two groups, as shown in Table 6. Based on the hypothesis that the SQ3R contributes to the emergence of repetitive learning behavior such as Prev, the SQ3R was applied in the experiment particularly to facilitate the occurrence of Prev behavior and achieve this purpose. For Prev, the mean and standard deviation were 24.43 and 26.07 for the control group and 60.39 and 35.71 for the experimental group. Moreover, the *t*-test result (t = -4.50, p < .01) showed that there was a significant difference between the two groups, implying that the SQ3R was able to promote the occurrence of Prev.

Analyzing the Read time variable alone, the difference between the two groups is apparent, and the *t*-test result (t = -2.35, p < .05) further verifies this conclusion, as shown in Table 6. However, the Read time variable was combined with the Read page variable for analysis. It is worth noting that the participants using the SQ3R took more time than those who did not, but after accounting for the Read page variable based on the *t*-test result (t = -4.83, p < .01), it was found that the former read nearly twice as efficiently as the latter. On average, the experimental group read approximately 3.01 pages per minute, compared with the control group's 1.59 pages per minute.

Second, we used the *t*-test to explore whether the SQ3R affects other reading behaviors. The results showed that there was a significant difference between the two groups in terms of Next (t = -3.42, p < .01), Memo (t = -3.16, p < .01), Underline (t = -3.62, p < .01), and Bookmark (t = -4.97, p < .01), though not for Highlight (t = .06, p > .05). These findings reveal that the SQ3R also promoted the occurrence of reading behaviors such as Next, Memo, Underline, and Bookmark. Regarding Highlight, no significant difference was found between the two groups.

Table 6. Descriptive data and *t*-test results for learning behavior

Variable	Group	*N*	Mean	*SD*	*t*
Prev	Control group	30	24.43	26.07	-4.50**
	Experiment group	31	60.39	35.71	
Read time	Control group	30	0:49:18	0:20:31	-2.35*
	Experiment group	31	1:00:30	0:16:30	
Read page	Control group	30	78.57	61.30	-4.83**
	Experiment group	31	182.32	101.96	
Underline	Control group	30	5.40	10.19	-3.62**
	Experiment group	31	37.52	48.16	
Highlight	Control group	30	5.30	18.22	-.060
	Experiment group	31	5.52	8.57	
Bookmark	Control group	30	.43	2.37	-4.97**
	Experiment group	31	6.90	6.82	
Memo	Control group	30	.00	.000	-3.16**
	Experiment group	31	2.19	3.85	
Next	Control group	30	43.00	28.42	-3.42**
	Experiment group	31	72.00	37.04	

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

Third, Figures 6 and 7 show learning patterns in terms of the time distribution of reading behaviors. The X-axis represents the time participants spent reading the material in the e-book system, and the Y-axis indicates the reading behaviors. The upper part of the two figures is the time distribution for each SQ3R step, where the time division between steps is based on the application guidance designed in the framework application step. Although there exists an obvious time division between various steps, this does not mean that all learning behaviors have a similar distribution to the preset learning steps.

Figure 6. Distribution of reading behaviors of the experimental group using the e-book system

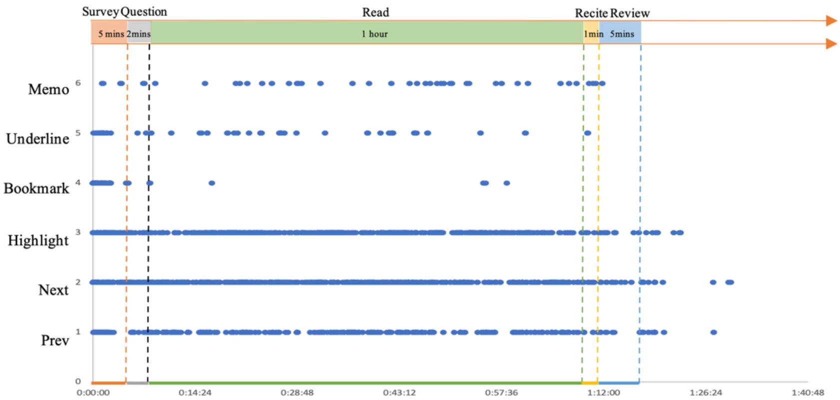


Figure 7. Distribution of reading behaviors of the control group using the e-book system

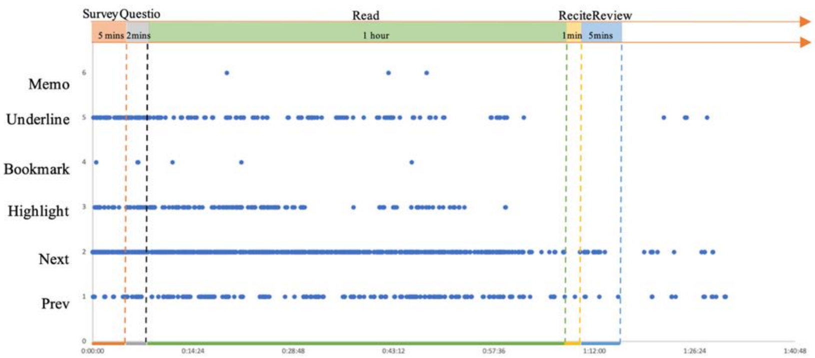


Figure 6 shows a rough distribution trend of reading behaviors representing three stages in terms of period: 0 min to 3 min, 3 min to 1 h 12 min, and 1 h 12 min to the end. Combining the five SQ3R divisions shows that in Stage 1, Prev, Underline, and Bookmark appear intensively in a short period, which indicates that the participants repeatedly read the material using the survey step, marking some knowledge points in a short time. Stage 2 accounts for most of the reading behavior. This stage might include the question and reading. The reading behaviors in Stage 3 primarily comprise Prev and Next; the other behaviors account for a smaller proportion. The time distribution of reading behaviors in the experimental group is consistent with the SQ3R, which suggests that participants spent 2 to 3 min on the first survey step, 2 min on the question step, and 5 min each on the reciting and review steps.

Figure 7 shows that the control group is divided into two stages, in which Prev and Next occur continuously and are the most frequent. Memo and Bookmark are less frequent, scattered, and irregular. In the first stage, Highlight and Underline appear frequently. Highlight has two distinct distributions, and it appears intensively and continuously for 30 min. Underline remains stable across several occurrences.

5.3. Analysis of cognitive load

A cognitive load post-questionnaire, which includes two test dimensions, mental load and mental effort, was employed to investigate the differences between the two groups, in terms of learning pressure and load on the participants. Because the mental load effect depends on the information being processed, which imposes a heavy cognitive load (Sweller, 1988), the first dimension focuses on the pressure caused by the amount of information the participants process. In addition, the second dimension carries out the mental effort test, which reflects the controlled consumption of psychological information processing resources in the cognitive process (Sweller, 1988; Hwang & Chang, 2011).

The *t*-test results in Table 7 show that for the mental load dimension, the mean and standard deviation were 11.91 and 3.25 for the control group and 11.17 and 2.35 for the experimental group. No significant difference was found between the two groups ($t = 1.01$, $p > .05$), implying that the SQ3R did not increase the pressure of information amount on participants. Regarding mental effort, no significant difference was found between the two groups ($t = 1.66$, $p > .05$). Therefore, the SQ3R did not exert additional pressure on participants in terms of mental load or mental effort. All participants in both groups had a moderate level of learning pressure, as indicated by the similar standard deviation concentrating at 3 for both groups.

Table 7. Descriptive data and *t*-test results for cognitive load

Variable	Group	*N*	Mean	*SD*	*t*
Mental load	Control group	30	11.91	3.25	1.01^{*}
	Experiment group	31	11.17	2.35	
Mental effort	Control group	30	10.47	3.27	1.66^{*}
	Experiment group	31	9.16	2.82	

Note. $^{*}p > .05$.

6. Discussion

This study explored whether the HAILA framework would affect the identification of optimal learning strategies. RQ1 investigated the effectiveness of the evidence-driven framework to offer guidance for transforming analysis results into the most suitable application strategy. For the application step performance, the SQ3R was obtained and optimally matched with the analysis results of the first experiment. The need for analysis result application has been proven by previous studies (Barry & Reschly, 2012); however, no previous study has provided concrete guidance for implementing the application. Inspired by HAI, evidence-driven education was incorporated into the design of the application step. Thus, this study explored the extent to which evidence-based education can facilitate the identification and transformation of analysis results into the optimal strategy. The experiment results show that evidence-driven education can sufficiently support application step design.

There remain downsides to the use of AI in LA, particularly related to algorithmic bias (Carter & Egliston, 2021). For example, decision-making based on AI analytics with unrepresentative datasets and algorithm design bias results in not only incomprehensible analysis results but also inappropriate or inapplicable result applications. Specifically, training AI algorithms on historical and complicated learning behavior data may reinforce the difficulty of understanding how it potentially undermines the learning behavior patterns. In that case, the HAILA framework contributes to increasing the interpretability of learning behavior patterns, not simply for exploring learning behavior patterns themselves. For example, the database on which AI algorithms are based is inevitably biased in terms of gender, family background, and ethnicity, resulting in the data itself containing bias (West et al., 2018). Moreover, concerns exist that AI analysis results can trigger artificial biases against learning behavior without an inappropriate or inapplicable application strategy, which reduces the trustworthiness of the AI analytics application. Therefore, the proposed framework with HAI consideration focuses on identifying the best application strategy that matches the analysis results. This function helps avoid artificial biases on learning behavior caused by inappropriate application strategies.

7. Conclusions

The HAILA framework is significantly effective in terms of analysis result application, and it highlights an implementable way to identify and apply the optimal application strategy, particularly concerning learning strategies. To verify the effectiveness of the modified framework and application strategy, two independent experiments were conducted. The results of the first experiment show that a learning strategy that best matched the analysis results was found through the application step in the framework. In addition, the findings of comparative experiments showed that students who applied the learning strategy achieved better learning results. This result is consistent with Li's et al. (2014) study showing that the SQ3R contributes to good learning achievement. However, it is unclear whether the SQ3R demands additional cognitive load. Moreover, a *t*-test showed that the experimental-group students who applied the learning strategy were not burdened with additional cognitive load, compared with the control-group students.

In contrast to the current analysis result application approaches, the HAILA framework consists of four steps, in which there are two design focuses. In particular, evidence-driven education is used to determine the optimal learning strategy, and the cognitive load test provides feedback on the application of the learning strategy. According to the results of the experiment, it was concluded that the SQ3R can improve academic performance by motivating the occurrence of repeated learning. Moreover, statistical results showed that the SQ3R helps increase the frequency of Prev learning behavior. In terms of the cognitive load test, there was no significant difference between the students who used the SQ3R and those who did not.

One of the purposes of HAI is to accurately identify at-risk students using AI algorithms and provide timely intervention that considers human education welfare. Some students are inevitably at risk of low academic performance; therefore, how to intervene promptly is a crucial problem. Based on previous studies, most research emphasizes identifying students who are at risk in terms of academic performance by analyzing learning behavior rather than applying learning strategies to overcome these risks. This study's primary contribution is that it succeeded in not only enhancing the trustworthiness of AI algorithms analysis results and verifying which factors contribute most to learning performance but also determining the optimal learning strategy for intervention in learning behavior to guarantee the effectiveness of AI algorithm analysis application.

Acknowledgement

This research was partially supported by the Grants-in-Aid for Scientific Research Nos. 21H00905 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan.

References

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

Bowers, A. J., & Sprott, R. (2012). Examining the multiple trajectories associated with dropping out of high school: A Growth mixture model analysis. *The Journal of Educational Research*, *105*(3), 176–195. https://doi.org/10.1080/00220671.2011.552075

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A New tool for a new era. *EDUCAUSE Review, 42*(4), 40–42.

Carter, M., & Egliston, B. (2021). What are the risks of virtual reality data? Learning analytics, algorithmic bias and a fantasy of perfect data. *New Media & Society, 21*(10), 1–20. https://doi.org/10.1177/14614448211012794

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A Reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318–331. https://doi.org/10.1504/ijtel.2012.051815

Choi, S. P. M., Lam, S. S., Li, K. C., & Wong, B. T. M. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Educational Technology & Society*, *21*(2), 273–290.

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In B. Wellman (Ed.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge: SESSION. Institutional perspectives* (pp. 134–138). ACM. https://doi.org/10.1145/2330601.2330636

Cook, B. G., Tankersley, M., Cook, L., & Landrum, T. J. (2008). Evidence-based practices in special education: Some practical considerations. *Intervention in School and Clinic, 44*(2), 69–75. https://doi.org/10.1177/1053451208321452

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, *47*(2), 108–121. https://doi.org/10.1111/1467-8527.00106

Dick, W., Carey, L., & Carey, J. O. (2015). *The Systematic design of instruction* (8th ed.). Pearson.

Dron, J., & Anderson, T. (2009). On the design of collective applications. In B. Randall (Ed.), *2009 International Conference on Computational Science and Engineering: Vol.4. Social intelligence and networking VII* (pp. 368–374). IEEE. https://doi.org/10.1109/CSE.2009.469

Elias, T. (2011). *Learning analytics: Definitions, processes and potential* (Computer Science; Corpus ID: 16906479) [Data set]. Semantic Scholar.

Flippo, R. F., & Bean, T. W. (Eds.). (2018). *Handbook of college reading and study strategy research* (3rd ed.). Madison Taylor & Francis Routledge. https://doi.org/10.4324/9781315629810

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, 22–32. https://doi.org/10.1016/j.compedu.2018.12.006

Hamoud, A. K., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(2), 26–31. https://doi.org/10.9781/ijimai.2018.02.004

Hanna, M. (2004). Data mining in the e-learning domain. *Campus-Wide Information Systems*, *21*(1), 29–34. https://doi.org/10.1108/10650740410512301

Hargreaves, D. H. (1997). In defence of research for evidence-based teaching: A Rejoinder to Martyn Hammersley. *British Educational Research Journal*, *23*(4), 405–419. https://doi.org/10.1080/0141192970230402

Huber, J. A. (2004). A Closer look at SQ3R. *Reading Improvement, 41*(2), 108–112.

Hwang, G. J., & Chang, H. F. (2011). A Formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education*, *56*(4), 1023–1031. https://doi.org/10.1016/j.compedu.2010.12.002

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. https://doi.org/10.1007/s13748-016-0094-0

Kuromiya, H., Majumdar, R., & Ogata, H. (2020). Fostering evidence-based education with learning analytics. *Educational Technology & Society, 23*(4), 14–29.

Li, L. Y., Fan, C. Y., Huang, D. W., & Chen, G. D. (2014). The effects of the e-book system with the reading guidance and the annotation map on the reading performance of college students. *Educational Technology & Society*, *17*(1), 320–331.

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R news, 2*(3), 18–22. https://www.researchgate.net/publication/228451484

Mesarić, J., & Šebalj, D. (2016). Decision trees for predicting the academic success of students. *Croatian Operational Research Review*, *7*(2), 367–388. https://doi.org/10.17535/crorr.2016.0025

Moran, D. J., & Malott, R. W. (2004). *Evidence-based educational methods.* Elsevier Academic Press.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71. https://doi.org/10.1207/s15326985ep3801_8

Park, H., & Kwon, H. C. (2011). Improved Gini-Index algorithm to correct feature-selection bias in text classification. *IEICE Transactions on Information and Systems*, *E94.D*(4), 855–865. https://doi.org/10.1587/transinf.e94.d.855

Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the Gini Index and Information Gain Criteria. *Annals of Mathematics and Artificial Intelligence*, *41*(1), 77–93. https://doi.org/10.1023/b:amai.0000018580.96245.c6

Rienties, B., Cross, S., & Zdrahal, Z. (2016). Implementing a learning analytics intervention and evaluation framework: What works? *Big Data and Learning Analytics in Higher Education*, 147–166. https://doi.org/10.1007/978-3-319-06520-5_10

Russell, R. S., & Taylor B. W. (2008). *Operations management* (4th ed.). Prentice Hall College Div. https://www.sciencedirect.com/book/9780750649957

Saks, K., Pedaste, M., & Rannastu, M. (2018). University teachers' and students' expectations on learning analytics. In M. Chang., N. S. Chen, R, Huang, K. Kinshuk, K. M. Moudgalya, S. Murthy, & D. G. Sampson (Eds.), *Proceedings of 18th International Conference on Advanced Learning Technologies* (pp. 183–187). IEEE. https://doi.org/10.1109/ICALT.2018.00050

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Siemens, G. (2013). Learning analytics. *American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tsai, S. C., Chen, C. H., Shiao, Y. T., Ciou, J. S., & Wu, T. N. (2020). Precision education with statistical learning and deep learning: A Case study in Taiwan. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00186-2

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The Current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

West, D., Tasir, Z., Luzeckyj, A., Si Na, K., Toohey, D., Abdullah, Z., Searle, B., Farhana Jumaat, N., & Price, R. (2018). Learning analytics experience among academics in Australia and Malaysia: A Comparison. *Australasian Journal of Educational Technology*, *34*(3). 122–139. https://doi.org/10.14742/ajet.3836

Xing, W., & Du, D. (2018). 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., Ogata, H., Matsui, T., & Chen, N. S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, *2*(1), 1–5. https://doi.org/10.1016/j.caeai.2021.100008

Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning: An International Journal, 10*(4), 455–468. https://doi.org/10.34105/j.kmel.2018.10.028

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2018). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A Human factor approach. *International Journal of Human-Computer Interaction*, *35*(4–5), 313–322. https://doi.org/10.1080/10447318.2018.1543077

Zhao, F., Hwang, G. J., & Yin, C. (2021). A Result confirmation-based learning behavior analysis framework for exploring the hidden reasons behind patterns and strategies. *Educational Technology & Society*, *24*(1), 138–151.