Zhao, F., Hwang, G.-J., & Yin, G. (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.

A Result Confirmation-based Learning Behavior Analysis Framework for Exploring the Hidden Reasons behind Patterns and Strategies

Fuzheng Zhao[1], Gwo-Jen Hwang[2] and Chengjiu Yin[1*]
[1]Kobe University, Japan // [2]National Taiwan University of Science and Technology, Taiwan //
zhaofz635@gmail.com // gjhwang.academic@gmail.com // yin@lion.kobe-u.ac.jp
[*]Corresponding author

ABSTRACT: Educational data mining and learning analytics have become a very important topic in the field of education technology. Many frameworks have been proposed for learning analytics which make it possible to identify learning behavior patterns or strategies. However, it is difficult to understand the reason why behavior patterns occur and why certain strategies are used. In other words, all of the existing frameworks lack an important step, that is, result confirmation. In this paper, we propose a Result Confirmation-based Learning Behavior Analysis (ReCoLBA) framework, which adds a result confirmation step for exploring the hidden reasons underlying the learning patterns and strategies. Using this ReCoLBA framework, a case study was conducted which analyzed e-book reading data. In the case study, we found that the students had a tendency to delete markers after adding them. Through an investigation, we found that the students did this because they could not grasp the learning emphasis. To apply this finding, we proposed a learning strategy whereby the teacher highlights the learning emphasis before students read the learning materials. An experiment was conducted to examine the effectiveness of this strategy, and we found that it could indeed help students achieve better results, reduce repetitive behaviors and save time. The framework was therefore shown to be effective.

Keywords: Learning analytics, Learning behavior pattern, Learning Analysis framework, Result confirmation

1. Introduction

In the last decade, technologies of educational data mining and learning analytics have made rapid progress (S. Baker, 2019). With the development of algorithm techniques, more scalable algorithms that can access an increasing number of computational resources have been proposed (Hashem, Yaqoob, Anuar, Mokhtar, Gani, & Khan, 2015). Moreover, there is greater availability of large amounts of fine-grained education data than before (Dietze, Siemens, Taibi & Drachsler, 2016). Despite it being easy to collect detailed learning data from multiple resources and to analyze them in order to provide educational suggestions or recommendations, the analysis results may not be sufficient to understand the critical information (Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-Gama, 2018).

In other words, through the use of some convenient data in a Web-based educational environment such as an e-book system, and the use of many mature analysis techniques such as sequence analysis, some learning behavior patterns can be easily found. However, it is not easy to understand the underlying reasons for behavior patterns. For example, Li, Uosaki, Ogata, Mouri, and Yin (2018) and Yin et al. (2019) found a backtrack reading behavior pattern which showed that certain students often return immediately to previous pages when they read e-textbooks; however, it is difficult to understand the reason for this particular behavior pattern. Therefore, it is very important to understand why the learning patterns or strategies behind behaviors occur, especially in terms of the complex social and cognitive processes involved.

In the philosophy of science, confirmation is an academic term related to the effect of evidence for hypotheses (Greco, Słowiński, & Szczęch, 2016). Based on that, result confirmation is viewed in this study as a step in which the reasons underlying behavior or strategies can be determined. Currently, the confirmation of the analysis results mainly depends on the judgment of the correct probability of the results obtained by the analysis algorithm. With this help, it can be inferred to what extent the correct analysis result has been obtained, in order to confirm it (Greco et al., 2016). However, this kind of confirmation emphasizes how relevant the learning behavior is to the analysis results, rather than the reasons underlying the behavior. Kennedy and Terry (2004) pointed out that cognitive components in user log data make it difficult to match a kind of behavioral pattern or strategy with a kind of behavior reason. There may be multiple reasons for the same behavior (Misanchuk & Schwier, 1992). Besides, it is noted that there has been limited analysis result confirmation in the field of learning analytics frameworks.

The main learning analytics frameworks proposed to date are: “Wisdom-Knowledge-Information-Data” (M. Baker, 2007), “Five Steps of Analytics” (Campbell & Oblinger, 2007), “Web Analytics Objectives” (Hendricks, Plantz, & Pritchard, 2008), "Collective Applications Model" (Dron & Anderson, 2009), "Processes of Learning Analytics" (Elias, 2011), and "Learning Analytics Processes" (Chatti, Dyckhoff, Schroede, & Thüs, 2012). It is evident that these learning analytics frameworks lack the indispensable confirmation step, which is a way to provide a structured process for the learning behavior analysis (Campbell & Oblinger, 2007). Based on previous research, it is easy to see that all of these frameworks usually consist of data collection, data processes, data analysis, and data application. While these frameworks have become more sophisticated, they all still lack the analysis result confirmation step. It is therefore difficult to identify the reasons behind the learners' behavior patterns and strategies. It also leads to ambiguous guidance for the application of education practice.

Based on the commonly used definition of learning analytics (Siemens & Baker, 2011), the primary function of a learning analytics framework can be summarized as understanding and optimizing learning and the corresponding contexts by the collection, measurement, analysis, and reporting of data. Notably, the functions of each component in the framework differ greatly. In terms of data analysis, it does not have the same functions as other components. For example, Yin and Hwang (2018) summarized the goals of data analysis as prediction, structure discovery, and relationship mining. With the advances in incorporating a variety of techniques, the requirements of the framework for educational practice have become increasingly precise, and have also been synchronized with the development of analytical techniques and learning theories. However, without confirming the analysis results, numerous obstacles have arisen in practice. In particular, the current research frameworks lack a crucial function of confirming the analysis results, which leads to an imbalance between practical requirements and framework functions; that is, the existing frameworks cannot meet the requirements for increasingly precise analysis results.

In this paper, we propose the Result Confirmation-based Learning Behavior Analysis (ReCoLBA) framework which includes a function to confirm the students' behavioral patterns and strategies. The ReCoLBA framework features the integration of confirming the reasons behind learning behavior into the existing frameworks. Its functional elements include data collection, data processing, data analysis, result confirmation, and result application.

This framework can also be applied to precision education. Recently, precision education has emerged as an important idea, and has been seen as having great potential for predicting which students are at risk, and providing timely interventions. On the basis of the extraction of the same philosophy between precision medicine and precision education, Lu et al. (2018) defined the objective of precision education as the improvement of diagnosis, prediction, treatment, and prevention of learning outcomes. However, relying on the existing learning analysis frameworks, we can only obtain the analysis of the recorded learning behavior, but cannot explain why this behavior occurs. Kennedy and Terry (2004) interpreted the reason for this as the lack of cognitive components in the electronic records of student activities. Through our proposed framework, we can not only determine which students are at risk, but also confirm the reasons for such behavior, so as to avoid a one-size-fits-all learning strategy.

Using the ReCoLBA framework, a case study was conducted by analyzing e-book reading data. In this ReCoLBA-based case study, we found that the students had a tendency to delete markers after adding them. Through an investigation, we found that they did this because they could not grasp the learning emphasis. To apply this finding, we proposed a learning strategy whereby the teacher highlights the learning emphasis on the e-textbooks before the students read the learning materials. At the same time two experiments were used, one to examine the usability of the framework, and the other to verify the effectiveness of the learning strategy proposed by the confirmed analysis results. Finally, The ReCoLBA-based case study illustrates that the framework is effective, and we found that the learning strategy could help students achieve better results and save time.

2. Literature review

2.1. The learning analytics framework

The term Educational Data Mining first appeared in a workshop in 2005, and then in 2008 the First International Conference on Educational Data Mining was held (Baker & Inventado, 2014). As a sister community to educational data mining, research on learning analytics and its frameworks followed. In 2007 M. Baker presented a framework, "wisdom-knowledge-information-data," calling it a "Knowledge Continuum." This framework emphasizes data processing in which knowledge is converted into a meaningful form (M. Baker, 2007). Compared with the above abstract learning analytic framework, Campbell, DeBlois, and Oblinger (2007) proposed the “Five Steps of Analytics”: capture, report, predict, act and refine, for a more simplified, practice-oriented structure procedure in the same year.

After the basic analytical framework had been constructed, coinciding with the rise of network analytic research, Hendricks et al. (2008) shifted the emphasis of the learning analytics framework study to the objectives of web analytics. From their perspective, the four operations of defining goals, measuring outcomes, using the resulting data, and sharing data for other stakeholders were identified as four objectives when using web analytics in education. These four objectives commonly constitute a learning analytics framework centered on the analysis objective.

With the recognition of the important role of social software in E-learning, Dron and Anderson (2009) noted that one kind of distinct dynamics has emerged in educational settings. It follows that group, network, and collective concepts gradually gained attention. Then, in 2009, they proposed a “Collective Applications Model” framework comprising five layers: selecting, capturing, aggregating, processing, and displaying. This framework was also classified into three cyclical phases: information gathering, information processing, and information presentation (Dron & Anderson, 2009). The cyclic structure with the head and tail connected was first put forward in this framework. In 2011, the definition of learning analytics was established at the 1st International Conference on Learning Analytics and Knowledge, followed by the learning analytics framework development in the theoretical dimension. Meanwhile, Elias provided a comprehensive learning analytics framework by summarizing the existing frameworks consisting of select, capture, aggregate & report, predict, use, refine, and share (Elias, 2011).

Despite the fact that these learning analytics frameworks were constructed in relation to the analysis process, there were still questions regarding combining them with the educational theory (Romero & Ventura, 2013). Chatti et al. (2012) introduced a successful learning theory, namely “Kolb’s Experiential Learning Cycle,” into the learning analytics for analysis framework. They also described a reference model for learning analytics, in light of an iterative cycle proposed as an older learning theory (Clow, 2012). This framework starts with data collection and processing, then goes into analytics and action, importantly, not ending with post-processing but entering a new cycle in a way that affects the next data collection. Similarly, in 2013, Siemens proposed a learning analytics cycle framework adopting a systems approach that includes seven components: collection, storage, data cleansing, data integration, analysis, visual presentation, and action (Siemens, 2013).

As a subfield of learning analytics, the learning behavior analysis framework basically follows the basic ideas of existing learning analytics frameworks. Although there are differences between the learning analysis frameworks mentioned above, a common feature is that each framework lacks an analysis result confirmation step. Therefore, we present the ReCoLBA framework, in which the result confirmation was added. Obviously, the distinctive function of this framework is to confirm why the behavior takes place and the strategies adopted.

2.2. Result confirmation

Confirmation is a term in the philosophy of science which is defined as the effect of evidence for a hypothesis. Here we define result confirmation as the analysis of the cause of the occurrence of behavior or strategies, which are obtained by learning analytics. The purposes of result confirmation are twofold: the first is the analysis algorithm confirmation, which is based on the probabilistic theory to infer the accuracy of the analysis results, whose evaluation metrics include support, confidence, correlation, and lift (Greco et al., 2016). With this help, it can be inferred to what extent the correct analysis result has been obtained, to complete the confirmation of the analysis result; the second is the analysis results confirmation. Generally, it is carried out using a pragmatic, mixed-methods approach (Phillips, 2006), and the analytical results are confirmed through self-reports, questionnaires, interviews, cases, and observations.

The result confirmation was designed to provide an explanation of the reasons behind the results. For example, the video click rate or duration is collected to predict future academic performance and dropout rates. If the analysis results are applied directly without confirmation, the positive effect of the results on practice is not guaranteed. Especially, solely relying on the click rate or duration is an analysis indicator, which is not causally related to academic performance, but only correlated. It has been proven that only using learning engagement to predict academic performance fails to accurately identify which student is at risk of dropping out. In fact, students who are usually considered as potentially giving up, are not necessarily those who drop out. Low engagement level has little to do with learning itself, and may be related to bad time management (Gourlay, 2017). Therefore, the real reasons behind these behavior results cannot be accurately discovered without confirmation, thus giving rise to the fuzzy specific guidance for stakeholders in education institutions.

2.3. Precision education

In the light of former US President Obama's 2015 State of the Union address where the precision medicine initiative was mentioned (Collins & Varmus, 2015; White House, Office of the Press Secretary, 2015), Hart widened the defining scope of precision medicine, including learning disabilities (LD) in the category, rather than it being limited to only biomedical diseases. Moreover, Hart (2016) argues that precision education would help researchers and practitioners understand the complex mechanisms underlying LD. Cook, Kilgus, and Burns (2018) defined precision education based on the best available evidence, and defined it as an approach to research and practice that is concerned with tailoring prevention and intervention practices to individuals. More recently, an investigation on how big data and artificial intelligence can be used to help universities more precisely understand student backgrounds was carried out (Tsai, Chen, Shiao, Ciou, & Wu, 2020). According to the results, the studies with the notion and science of precision education may enable universities to provide interventions to students for course selection and competence growth (Hart, 2016).

3. Result confirmation-based learning behavior analysis framework

As shown in Figure 1, the ReCoLBA framework, which consists of five main steps: data collection, data processing, data analysis, result confirmation, and result application, has a cyclic and iterative structure. Based on the two frameworks presented by Chatti et al. (2012) and Siemens (2013) and Kolb's Experiential Learning Cycle learning theory, we propose a Result Confirmation-based Learning Behavior Analysis (ReCoLBA) framework in this study.

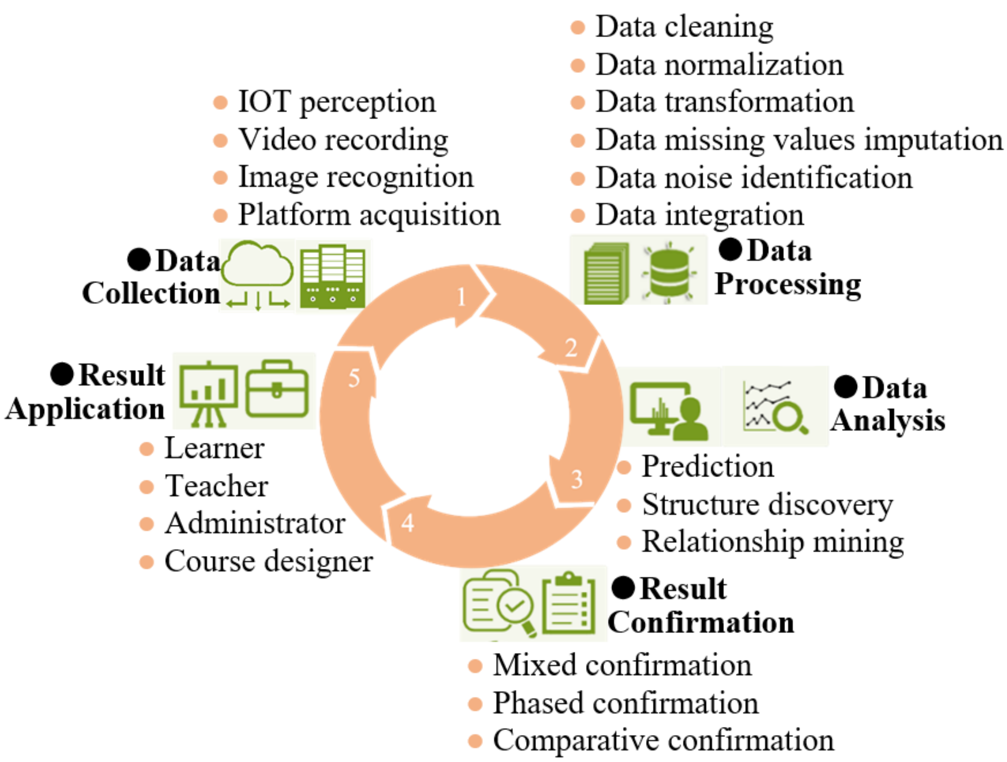

Figure 1. The ReCoLBA framework

By comparing and combining the aforementioned frameworks adopted by the ReCoLBA, while clarifying the differences and similarities between them, we highlight the innovative functions of this framework in Table 1. The most important functions of identifying and confirming the reasons underlying the learning behavior are obvious. Moreover, an important education theory, Kolb's Experiential Learning Theory, was introduced into the ReCoLBA framework. The integration of that theory provides a cyclic and iterative structure for this framework. Subsequently, a specific introduction of each step will be provided.

Table 1. The differences and similarities between the reference frameworks and the ReCoLBA framework

Name	ReCoLBA framework	Learning Analytics framework (Chatti et al., 2012)	Processes of Learning Analytics framework (Siemens, 2013)
Steps	Data collection	Data collection	Collection Storage
	Data processing	Pre-processing	Data cleaning Data Integration
	Data analysis	Analytics	Representation and visualization
	Result confirmation	None	None
	Result application	Action and Post-processing	Action

3.1. Data collection

The first primary function of this framework is to collect data from different education environments, with the main data collecting methods, such as IOT perception, video recording, image recognition, and platform acquisition. According to the report "SEDCAR, Standards for Education Data Collection and Reporting" (National Center for Education Statistics, 1991), data in educational settings or systems can be collected and analyzed by means of record extraction, surveys (mail, telephone, face-to-face), observations, experiments, and secondary data analysis. Usually, the collection methods can be divided into two categories: designed experiments and the observational approach. The former means that collectors control the data generation conditions. The latter means that collectors do not participate in the data generation process (Kantardzic, 2011). At present, data collection technologies include the following four main technical categories (Wong, 2017): Internet-of-Things (wearable devices), video recording technology (video broadcasting), learning platform acquisition technology (log data), and image recognition technology (eye-tracking).

3.2. Data processing

The second data processing step includes data cleaning, data normalization, data transformation, data missing values imputation, data noise identification, and data integration (Romero & Ventura, 2010a; García, Ramírez-Gallego, Luengo, Benítez, & Herrera, 2016). Data processing is a step that transforms raw data into a useful and efficient format (Chakrabarty, Mannan, & Cagin, 2015). The main tasks of data processing are to retrieve inaccurate records in the data set, identify incorrect or irrelevant records in the data set, and manipulate the collected data by deleting, modifying, and replacing (Wu, Zhu, Wu, & Ding, 2013).

3.3. Data analysis

The following step is data analysis, which consists of the three analysis goals of prediction, structure discovery, and relationship mining. Data analysis is an operation which is guided by certain research purposes, such as prediction, structure discovery, and relationship mining (Yin & Hwang, 2018). Each particular educational problem has its own specific objectives, so the existing analytical methods and techniques cannot be directly applied to the analysis of such data (Romero & Ventura, 2013). In other words, the analysis methods are not categorized into special research areas in the education field. The data analysis methods are divided into 11 methods according to three categories: prediction, structure discovery, and relationship mining (Yin & Hwang, 2018), as shown in Table 2.

Table 2. Goals and methods of data analysis

Goals	Prediction	Structure discovery	Relationship mining
Methods	Classification	Clustering	Association rule mining
	Regression	Factor analysis	Correlation mining
	Latent knowledge	Knowledge inference	Sequential pattern mining
	Estimation	Network Analysis	Causal data mining

3.4. Result confirmation

The obvious function point of the framework which differs from any others is the result confirmation, which is made up of mixed confirmation, phased confirmation, and comparative confirmation. This study reviewed previous findings on the learning analytics frameworks, and found that they were lacking a results confirmation step. To address the problems of confirming the analysis results, four representative research papers were identified related to how to conduct a study to confirm the analysis results. For example, to evaluate the online materials in the context of the unit of study, Phillips, Baudains, and Van (2002) confirmed existing results by using learning logs recorded in the WebCT site for students' approaches. In addition to quantitative approaches to confirming learning outcomes using statistical methods, such as comparison of the previous research in early years, an investigation was employed to confirm the situation about completing laboratories, ongoing study, and surface learning. Only relying on the understanding of the fact that the surface analysis using data-based learning analytics is insufficient to identify student learning behavior, a case study of how qualitative data provide rich information to confirm the analysis results was carried out by Phillips (2006), and a comparative confirmation using the data generated from the study subjects in different periods was used to discover the changing laws (Kennedy & Terry, 2004; Li et al., 2018). In response to the insufficient interpretation of data analysis for the learning behavior, they interviewed the learners to confirm the potential interactions after the data analysis. By carrying out investigations, this approach with phased confirmation steps can identify the potential reasons underlying the existing analysis, such as why students repeat the learning behavior.

Based on the specific confirmation methods presented in these four papers, this study summarizes them into the three categories of mixed confirmation, comparative confirmation, and phased confirmation.

- Mixed confirmation. This is a method of confirming the analysis results through a combination of quantitative and qualitative studies. It not only includes using evaluation metrics of the analysis algorithm, but also involves some qualitative methods such as documentation, staff interviews, observation of students in laboratory classes, and a student survey (Phillips et al., 2002).
- Comparative confirmation. There are two confirmation dimensions of the comparative confirmation method. The first is the time dimension. The comparison of the similarities and differences is used to discover the changing laws, by means of collecting the data generated from the study subjects in different periods. Data collection intervals are chosen, based on semesters or school years. The second is the spatial dimension. For example, students' learning behavior data from different schools are comparatively analyzed for results confirmation (Phillips, 2006).
- Phased confirmation. The results confirmation involves two analytic stages. The statistical analysis is usually adopted in the first stage. If the variables of interest had occurred, or the analysis results had abnormal values, then they would move to the second stage. The focuses on interesting variables or abnormal values will be analyzed again. The confirming method includes interviews and questionnaires (Kennedy & Terry, 2004; Li et al., 2018).

3.5. Result application

The last one is the result application step which involves four stakeholders: learner, teacher, administrator and course designer. It is noteworthy that the final goal of the ReCoLBA framework is to apply the analysis results to educational practice. Different roles can derive different benefits from the confirmed results of learning analytics. For example, the confirmed analysis results could help students share their learning experience, help teachers master students' learning behavior and provide them with timely support, help manager administrators to evaluate teachers and students, and help course designers to improve the course content and instructional materials according to the confirmed analysis results (Yin & Hwang, 2018).

4. Experimental design

Using the ReCoLBA framework, we designed a case study including two experiments, one aiming to examine the usability of the framework, and the other focusing on verifying the effectiveness of a learning strategy proposed by the confirmed analysis results.

4.1. Experiment for verifying the usability of the ReCoLBA framework

To examine the usability of the proposed framework, a ReCoLBA-based case study was conducted. To this end, 48 participants were recruited and assigned to read academic papers using an e-book system. There are five parts in this experiment, namely using the e-book system to collect students' reading behavior, performing five steps of data processing, adopting behavioral sequential analysis for data analysis, utilizing a mixed confirmation method to confirm the analysis results, and applying a learning strategy by an intervention.

4.1.1. Data collection in the case study

An e-book system was developed corresponding with the functions of the ReCoLBA framework, to evaluate the framework usability. As shown in Figure 2, this system can collect the students' reading behavior when reading learning materials, such as (1) page-turning, (2) zoom+/-, (3) writing a memo, (4) adding or deleting an underline, and (5) adding or deleting a highlight. Most importantly, all reading behavior actions were recorded in the form of reading action logs. Moreover, the e-book system can provide automatically coded action logs as one of the convenient functions to reduce subsequent data processing. In this way, the usual manual coding process completed by programmers can be avoided.

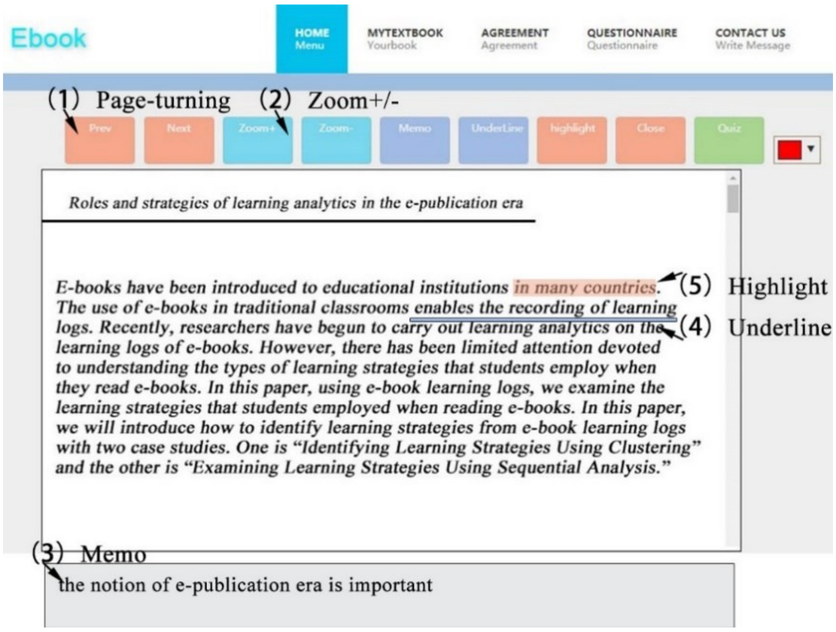

Figure 2. The E-book system interface

A total of 4,748 records were collected from 60 graduate students, who were asked to complete reading academic papers in the e-book system within 90 mins in this experiment. Their reading behavior data, which consisted of user ID, operation name, page number, and action time, were stored in the database for analyzing their learning behavior patterns or strategies, as shown in Table 3.

Table 3. Sample action log

User ID	Operation name	Page number	Action time
Student 1	NEXT page:0	19	2020/7/10 12:05:18
Student 2	PREV page:8	16	2020/7/10 12:07:31
Student 3	ADD HL page:4	16	2020/7/10 12:36:54

4.1.2. Data processing in the case study

This study summarizes data processing as the following four points. The first is data transformation. Through the statistical processing, the sum of the students' reading behavior including adding or deleting underlines, adding or deleting highlights, and adding or deleting memos, was respectively counted. The second point is data cleaning. If the following learning phenomenon or behavior occurred during the learning activity, it was viewed as an invalid record. For example, if the longest duration between two learning behaviors exceeded 20 minutes, then the record was invalid as it indicated that no reading activities had occurred because the student did not conduct any learning behavior within 20 minutes. Besides, incomplete records were filtered. The third point is missing values imputation and noise identification. The valid sample size changed from 60 to 47 by identifying and discarding missing values data and noisy data. The fourth point is data normalization. There were some data completed by those who had an invalid preview, for example, some students who completed the preview of the lesson (read the learning content before class) less than 3 minutes before the class. In that case, the preview of the lesson was viewed as invalid. The final point is data integration. The objective of data processing is mainly from the e-book system, so that the sole data generating source avoids the need to integrate data from different data sources, which makes it unnecessary to adopt data transformation and integration in the data processing.

4.1.3. Data analysis in the case study

In the case study, behavioral sequential analysis was adopted to gain a detailed understanding of progressive learning behavioral patterns (Bakeman & Gottman, 1997; Hou, 2012). The progressive learning behavior patterns were obtained by sequential analysis, as shown in Figure 3. The squares represent learning behavior, and the arrow-lines and numbers represent the direction and extent to which the behavior is associated with other behaviors. Rounded curves represent the association of the behavior with itself. The result is significant at the < .05 level when the *z*-value is greater than 1.96 (Bakeman & Gottman, 1997; Hou, 2012).

In Figure 3, it is obvious that "HIGHLIGHT" is mutually associated with "DEL HIGHLIGHT" and "DEL UNDERLINE," and "UNDERLINE" is mutually associated with "DEL HIGHLIGHT" and "UNDERLINE." Meanwhile, "HIGHLIGHT," "DEL HIGHLIGHT," "DEL UNDERLINE," and "UNDERLINE" are respectively associated with themselves. "BOOKMARKER" is mutually associated with "DEL BOOKMARKER." Finally, "BOOKMARKER" has a one-way association with "UNDERLINE" and "MEMO." The association between this repetitive behavior was confirmed to be significant, using the Z-score binomial test parameters in the sequential analysis method. However, this is not sufficient to understand why these patterns and strategies occurred, and especially to understand why some students exhibit some repetitive behavior of the same operation. For example, some students deleted UNDERLINE after adding an UNDERLINE, and some deleted the HIGHLIGHT after adding a HIGHLIGHT.

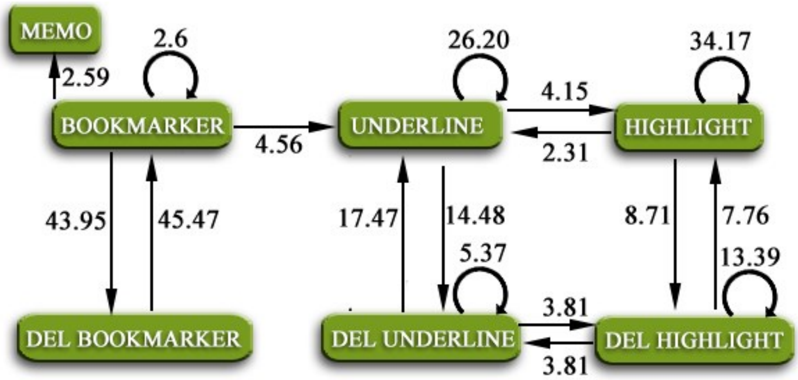

Figure 3. The students' progressive learning behavioral patterns

4.1.4. Resulting confirmation in the case study

We identified the repetitive learning behavior patterns as "Deleting a HIGHLIGHT after adding the HIGHLIGHT," "After adding an UNDERLINE, delete the UNDERLINE." However, it was difficult to understand the reason why these behavior patterns happened. Therefore, it was necessary to conduct the result confirmation step. To this end, an investigation system was used to investigate the participants who had repetitive learning behavior patterns, as shown in Figure 4. The participants were investigated to answer the questions that matched their behavior such as, "Why did you adopt "After adding a HIGHLIGHT, delete the HIGHLIGHT," and why did you adopt "After adding an UNDERLINE, delete the UNDERLINE." Samples of extracts from their answers are shown in Table 4.

In this investigation, only if common results were gained by two researchers was the coding of the students' answers accepted. The notion for the data analysis of up to 24 participants was extracted, and the reference basis is strong. After the investigation, it was confirmed that it was difficult for some students to identify which were the keywords while reading. Other students often read repeatedly to understand the main idea of the papers because they had difficulty finding the most important content. It was concluded that the repetitive operation of markers happened because they could not correctly identify the learning emphasis. Besides, the marking method, such as underline and highlight, is beneficial for students to mark the important content, and these choices are mainly related to personal interest.

Figure 4. The investigation system interface

Table 4. Samples of participants' answer content

Participant	Answer
Student 1	When I first read it, I highlighted the content that I thought was important. However, going back to read it a second time, I realized that it was not the most important content anymore, so I decided to delete it.
Student 2	It is useful to deepen my understanding of the paper.
Student 3	When I read the sentences, from my perspective, it is more suitable for the question.
Student 4	I subsequently added an underline after the highlight, which is my way of distinguishing the emphasis.
Student 5	That is a mistake in the highlight position owing to not being proficient in operating the platform.

4.1.5. Resulting application in the case study

Through the ReColBA-based case study, it was found that the reason why students exhibit the repetitive behavior of the same operation is that they could not find the learning emphasis. The e-textbooks lack any marking of the learning emphasis, which is not conducive to students' understanding. Therefore, we proposed a learning strategy that marked the learning emphasis in the e-textbooks using underlines and highlights, as shown in Figure 5.

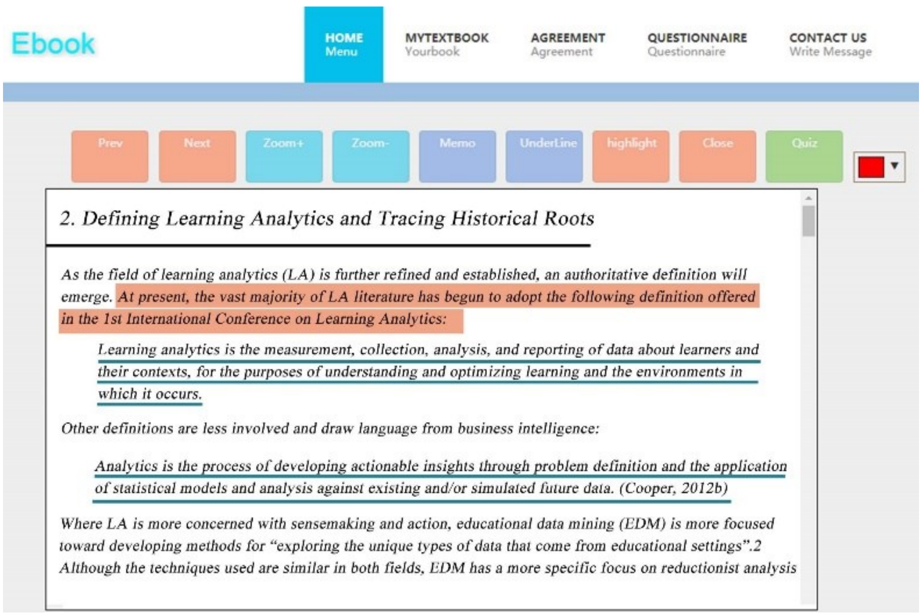

Figure 5. The e-textbook with marking of the learning emphasis

4.2. Experiment for verifying the learning effectiveness of the ReCoLBA framework

To verify the learning effectiveness of the learning strategy discovered by the framework, we conducted a verification experiment. The specific steps are as follows.

4.2.1. Participants

A total of 80 graduate students from a university's graduate school were recruited to participate in this experiment. As the experimental group, 43 students read an e-textbook via the e-book system, using the marking learning emphasis method. Another 37 students in the control group adopted the same experiment conditions, except that they did not use the marking learning emphasis technique.

4.2.2. Measuring methods

The measuring methods consisted of a pre-test, a post-test, and an interview. Besides, the frequency and duration of learning behavior which were automatically recorded by the e-book system were used to evaluate the effectiveness of the learning strategy. Firstly, the pre-test was to evaluate whether both groups had equivalent prior knowledge of the upcoming learning content. It included 10 multiple-choice items. Secondly, the post-test was to measure whether the marking learning emphasis method was helpful for students' learning achievement. It also included 10 multiple-choice items, which were related to the core learning emphasis of the article. The total score was 100 in both the pre-test and post-test. Thirdly, a 30-minute semi-structured interview was conducted for the experimental group which was recorded verbatim. The following questions were asked: What do you think of this learning method? What is the difference between this way of marking the key points of the paper and your previous learning method? Two researchers were invited to analyze the interview data, and to determine the participants' attitudes towards the proposed learning strategies through the extracted core keywords from the interview content.

4.2.3. Experimental procedure

The learning emphasis aimed to help students understand the definition, historical development, and application fields of Learning Analytics, which is a unit of an Education Technology course in the university.

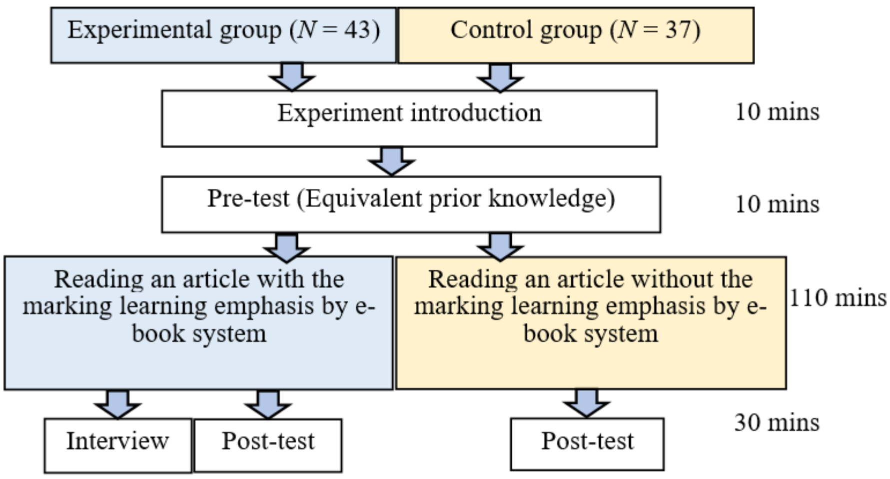

Figure 6. Experiment design diagram

As shown in Figure 6, before the experiment, the students were given a pre-test to test their prior knowledge of the learning emphasis. Subsequently, we introduced the learning emphasis and e-book system, then conducted a 110-minute learning activity. Both the experimental group and the control group were required to use the e-book system functions at will, such as page-turning, underlining, highlighting, and making memos. Unlike the control group, the e-textbook for the experiment group was already underlined and highlighted, to indicate the learning emphasis. After the learning activity, a post-test was conducted, and the experimental group students were asked to take part in an interview.

5. Experiment results

First, we analyzed the participants' learning achievement. The effectiveness of the proposed method was examined to ensure whether this approach could improve the students' learning performance. Based on the pre-test, the standard deviations and mean values were 20.471 and 60.00 for the control group, and 15.596 and 78.92 for the experimental group. According to the *t*-test result ($t = 2.897$, $p > .05$), as shown in Table 5, there was no significant difference between the two groups. It was therefore found that they had equivalent prior knowledge before the learning activity.

Table 5. Descriptive data and *t*-test result of the pre-test results

Variable	Group	*N*	Mean	*SD*	*t*
Pre-test	Experimental group	43	60.00	20.471	2.897
	Control group	37	78.92	15.596	

The post-test was taken after the learning activity. We used the one-way analysis of covariance (ANCOVA) to evaluate the learning achievement of the two groups, setting the groups as a fixed factor, the pre-test scores as the covariate, and the post-test scores as the dependent variable. The data met the ANCOVA requirements. The Levene's test of equality of error variances ($F = 1.409$, $p > .05$) indicated that the assumption of regression homogeneity was followed. The results of the ANCOVA ($F = 18.424$, $p < .01$) in Table 6 determined that the experimental group achieved better results. It was concluded that the marking learning emphasis method was beneficial for the students to understand the learning emphasis.

Table 6. Descriptive data and one-way ANCOVA result of the post-test results

Variable	Group	*N*	Mean	*SD*	Adjusted mean	*F*
Post-test	Experimental group	43	88.60	13.378	89.74	18.424^{*}
	Control group	37	77.62	11.526	76.29	

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

Second, we analyzed the system reading time. To better understand the effectiveness of this method, the ANCOVA was also used to analyze the reading time of the two groups. The Levene's test of equality of error variances ($F = 1.409$, $p > .05$) was not violated. From the ANCOVA results in Table 7, it can be seen that the means and standard deviations of the experimental group are 0:41:06 and 0:21:21, and for the control group they are 1:12:27 and 0:26:26. There was a significant difference ($F = 41.731$, $p < .01$) between the two groups. It was clear that the experimental learners spent less system reading time finishing the learning task than those who used the conventional method. It was found that it was helpful to save time compared with the students who did not use the marking learning emphasis method.

Table 7. Descriptive data and one-way ANCOVA result of reading time

Variable	Group	*N*	Mean	*SD*	Adjusted mean	*F*
Reading time	Experimental group	43	0:41:06	0:21:21	0:38:12	41.731^{*}
	Control group	37	1:12:27	0:26:26	1:16:32	

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

Third, we analyzed the results of the Highlight and Underline operations. To examine the effectiveness of the intervention by marking the learning emphasis, the deleting rate of highlight and underline operations was viewed as a dependent variable, to test whether there were significant changes in the repetitive behaviors. In terms of the deleting rate of the highlight results, it was found that the standard deviation and mean values were 0.214 and 0.100 for the control group, and 0.040 and 0.010 for the experimental group. The *t*-test result ($t = 0.016$, $p < .05$), as shown in Table 8, indicated a significant difference between the two groups. For the underline results, it can be seen that the mean values and standard deviations of the experimental group are 0.011and 0.053, and for the control group they are 0.071 and 0.158. There was a significant difference ($t = 0.032$, $p < .05$) between the two groups. It is concluded that a significantly decreasing tendency of the deleting rate occurs, which illustrates that the repetitive behaviors were reduced.

Table 8. Descriptive data and *t*-test result of the deleting rate of the Highlight and Underline results

Variable	Group	*N*	Mean	*SD*	*t*
Highlight	Experimental group	43	0.010	0.040	0.016
	Control group	37	0.100	0.214	
Underline	Experimental group	43	0.011	0.053	0.032
	Control group	37	0.071	0.158	

Fourth, we analyzed the interview data. After the experiment, an interview was conducted on the topic of how to evaluate the learning methods, and 43 interview transcripts were obtained. It was concluded that 42 participants gave positive evaluations and one participant had a neutral attitude. Two dimensions constituted a positive evaluation, namely Knowledge comprehensibility (27 participants) and Reading convenience (32 participants). In all, 27 participants used the key items "Good for memorizing content" (14 participants), "Strong visual guidance" (4 participants), "Helps me understand the content" (6 participants) and "Convenient for future work" (3 participants) to describe what benefits it brought. In terms of reading convenience, 32 participants shared their positive evaluations. Four participants mentioned that this method could increase reading convenience, 18 argued that they could read better with the help of the reading emphasis marked by underlining or highlighting, and 10 maintained that the efficiency gains and time-saving were the most important benefits of the method.

6. Conclusions

The existing research clearly indicates that the analysis results have the hidden probability of creating misunderstandings of students' behavior in practice, especially in the case of lacking result confirmation. Owing to the cognitive element and multiple interpretations of learning behavior, result confirmation requires not only providing the patterns or strategies for learning behavior, but also identifying the reasons behind their occurrence, and only by doing so will the analysis results be better applied in practice.

A case study was then conducted to verify the usability of the ReCoLBA framework. Based on the findings concluded from the application experiment, the ReCoLBA framework has been successfully verified. It can help identify why certain learning behaviors occur and the strategies adopted, and avoids applying analysis results without confirmation. The investigation in the result confirmation step was different from that usually employed in the domain of human science. This investigation method is combined with the analysis technique, and the subjects that needed to be studied were based on the primary analysis results, which was deduced from the data.

The confirming function designed in the ReCoLBA framework enables researchers to have an opportunity to access the reasons underlying the learning behavior. Through the application of the framework, it was found that the reason for the students repeatedly adding and deleting underlines or highlights was the lack of learning emphasis. Based on the findings, an e-textbook with the learning emphasis marked with underlines and highlights was designed and developed. To verify the effectiveness of the revised learning method of previously giving the learning emphasis, an experiment was conducted. From the results of the *t*-test, one-way ANCOVA, and interviews, it was obvious that the marking learning emphasis method could help students improve their learning achievement and save time when using the revised e-textbook. Through verification, we found that our framework can help teachers discover which students are at risk of failing the course, and effectively confirm the results of the behavioral analysis. Specific interference strategies are then proposed for confirmed learning behavior, which is consistent with the goal of precision education.

Through a literature review, it was found that there are three kinds of methods to confirm the analysis results, that is, mixed confirmation, phased confirmation, and comparative confirmation. We subsequently integrated these three methods, which can be used separately for confirming analysis results, as a new step in the ReCoLBA framework. In that case, there is more space left to discuss the other confirmation method. It is noted that the ReCoLBA framework can help teachers to offer at-risk students a precise intervention by using a precise guiding strategy that is not only limited to teachers, but is also suitable for the different stakeholders in education, such as the precise management by administrators, precise self-regulation by learners, and precise course design by course designers.

Acknowledgement

Part of this research work was supported by the Grant-in-Aid for Scientific Research No.18H04125 and No. 16H06304 from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) in Japan.

References

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An Introduction to sequential analysis*. Cambridge, UK: Cambridge university press.

Baker, B. M. (2007). *A Conceptual framework for making knowledge actionable through capital formation* (Unpublished doctoral dissertation). University of Maryland University College.

Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *JEDM Journal of Educational Data Mining*, *11*(1), 1-17.

Baker, R. S., & Inventado, P. S. (2014). Chapter 4: Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics: From Research to Practice* (pp. 61-75). New York, NY: The Springer Science+Business Media Company.

Campbell, J. P., & Oblinger, D. G. (2007). Academic analytics. *EDUCAUSE white paper*. Retrieved from https://net.educause.edu/ir/library/pdf/PUB6101.pdf

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A New tool for a new era. *EDUCAUSE Review*, *42*(4), 40-57.

Chakrabarty, A., Mannan, S., & Cagin, T. (2015). *Multiscale modeling for process safety applications*. Waltham, MA: The Butterworth-Heinemann Company.

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), 318-331.

Clow, D. (2012). The Learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134-138). New York, NY: Association for Computing Machinery.

Collins, F. S., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine, 372*, 793–795.

Conde, M. Á., & Hernández-García, Á. (2015). Learning analytics for educational decision making. *Computers in Human Behavior*, *47*(7), 1-3.

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology, 66*(11), 4-10.

Dietze, S., Siemens, G., Taibi, D., & Drachsler, H. (2016). Datasets for learning analytics. *Journal of Learning Analytics*, *3*(2), 307-311.

Dron, J., & Anderson, T. (2009). On the design of collective applications. In *Proceedings of the 2009 International Conference on Computational Science and Engineering* (pp. 368-374). Vancouver, BC: IEEE.

Elias, T. (2011). Learning Analytics: Definitions, Processes and Potential. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: Methods and prospects. *Big Data Analytics, 1*(1), 9-31.

Gourlay, L. (2017). Student engagement, "learnification" and the sociomaterial: Critical perspectives on higher education policy. *Higher Education Policy*, *30*(1), 23-34.

Greco, S., Słowiński, R., & Szczęch, I. (2016). Measures of rule interestingness in various perspectives of confirmation. *Information Sciences*, *346*(1), 216-235.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The Rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*(7), 98-115.

Hart, S. A. (2016). Precision education initiative: Moving toward personalized education. *Mind, Brain, and Education, 10*(4), 209-211.

Hendricks, M., Plantz, M. C., & Pritchard, K. J. (2008). Measuring outcomes of United Way–funded programs: Expectations and reality. *New Directions for Evaluation*, *119*(9), 13-35.

Hou, H. T. (2012). Exploring the behavioral patterns of learners in an educational massively multiple online role-playing game (MMORPG). *Computers & Education*, *58*(4), 1225-1233.

Kantardzic, M. (2011). *Data mining concepts models methods and algorithms*. Hoboken, NJ: Wiley-IEEE Press.

Kennedy, G. E., & Judd, T. S. (2004). Making sense of audit trail data. *Australasian Journal of Educational Technology*, *20*(1), 18-32.

Li, L., Uosaki, N., Ogata, H., Mouri, K., Yin, C. (2018) Analysis of Behavior Sequences of Students by Using Learning Logs of Digital Books. In *Proceedings of the 26th International Conference on Computers in Education* (pp. 367–376). Manila, Philippines: APSCE.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., 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*(3), 179-196.

Misanchuk, E. R., & Schwier, R. A. (1992). Representing interactive multimedia and hypermedia audit trails. *Journal of Educational Multimedia and Hypermedia*, *1*(3), 35-72.

National Center for Education Statistics. (1991). SEDCAR (*Standards for Education Data Collection and Reporting*). Rockville, MD: Westat, Inc.

Phillips, R. A. (2006). Tools used in Learning Management Systems: Analysis of WebCT usage logs. In L. Markauskaite, P. Goodyear & P. Reimann (Eds.), *Proceedings of the 23rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Who's Learning? Whose Technology?* (pp. 663-673). Sydney, Australia: Sydney University Press.

Phillips, R. A., Baudains, C., & Van Keulen, M. (2002, December). An Evaluation of student learning in a web-supported unit on plant diversity. In A. Williamson, C. Gunn, A. Young & T. Clear (Eds.), *Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 525-534). Auckland, NZ: UNITEC Institute of Technology.

Romero, C., & Ventura, S. (2010). Educational data mining: A Review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601-618.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12-27.

Siemens, G. (2013). Learning analytics: The Emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380-1400.

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*(4), 1-13.

White House, Office of the Press Secretary. (2015). *Fact sheet: President Obama's Precision Medicine Initiative*. Retrieved from https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative.

Wong, B. T. M. (2017). Learning analytics in higher education: An Analysis of case studies. *Asian Association of Open Universities Journal*, *12*(1), 21-40.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. *IEEE transactions on knowledge and data engineering*, *26*(1), 97-107.

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.

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). 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), 313-322.