

Article

Integrating Business Analytics in Educational Decision-Making: A Multifaceted Approach to Enhance Learning Outcomes in EFL Contexts

Minsu Cho , Jiyeon Kim, Juhyeon Kim and Kyudong Park * 

Department of Artificial Intelligence Applications, Kwangwoon University, Seoul 01897, Republic of Korea; mcho@kw.ac.kr (M.C.);

* Correspondence: kdpark@kw.ac.kr; Tel.: +82-10-8972-0610

Abstract: This study introduces a framework that integrates business analytics into educational decision-making to improve learner engagement and performance in Massive Open Online Courses (MOOCs), focusing on learning environments in English as a Foreign Language (EFL). By examining three specific research questions, this paper delineates patterns in learner engagement, evaluates factors that affect these patterns, and examines the relationship between these factors and educational outcomes. The study provides an empirical analysis that elucidates the connection between learner behaviors and learning outcomes by employing machine learning, process mining, and statistical methods such as hierarchical clustering, process discovery, and the Mann–Kendall test. The analysis determines that learning patterns, characterized as single-phase or multi-phase, repetitive or non-repetitive, and sequential or self-regulated, are more closely associated with the nature of the educational content—such as books, series, or reading levels—than learner characteristics. Furthermore, it has been observed that learners exhibiting self-regulated learning patterns tend to achieve superior academic outcomes. The findings advocate for integrating analytics in educational practices, offer strategic insights for educational enhancements, and propose a new perspective on the connection between learner behavior and educational success.



Citation: Cho, M.; Kim, J.; Kim, J.; Park, K. Integrating Business Analytics in Educational Decision-Making: A Multifaceted Approach to Enhance Learning Outcomes in EFL Contexts.

Mathematics **2024**, *12*, 620. <https://doi.org/10.3390/math12050620>

Academic Editors: Ou Liu, Heng Tang and Jay Jahangiri

Received: 6 January 2024

Revised: 7 February 2024

Accepted: 18 February 2024

Published: 20 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: learning analytics; business analytics; process mining; machine learning; learning patterns

MSC: 97C70

1. Introduction

In the realm of business operations, decision-making processes hold paramount significance as they constitute a fundamental facet of effective business management [1,2]. Within this context, decision-makers increasingly rely on empirical evidence from data-driven analyses, i.e., business analytics, as the bedrock for their choices rather than intuitive judgment or fragmented information [1,2]. Hence, developing a framework that integrates business analytics into decision-making cycles is essential across many domains. Such a framework ensures that decisions are continually informed and refined by the latest data, leading to more strategic and effective business outcomes [3,4].

Within an educational setting, the primary objective revolves around providing student-centric instructional content and facilitating sound decision-making for learners [5–7]. Learning analytics thus emerges as a pivotal tool within the educational business model [6]. The growth of diverse online educational environments has expanded course accessibility but also brought challenges, including high dropout rates, insufficient guidance, and limited intervention options [8]. In this context, it is essential to leverage data analytics to gain insights into the learning environments of these students through data-driven analyses. Collaborating with this knowledge, educators can proactively offer tailored support

designed to meet the specific needs of students, thereby improving their engagement before they discontinue the course [9].

Figure 1 delineates the overview of decision-making augmented by learning analytics. Contemporary Learning Management Systems (LMS) are critical in orchestrating and evaluating the various facets of the educational process. In addition to encompassing the planning, execution, and evaluation of learning processes, these systems also serve as repositories for the learning data pertaining to learners' interactions and activities. In essence, the collected learning data represent a valuable resource that, upon analysis, yields meaningful insights into learners' behavioral patterns and characteristics. Subsequently, these insights can be translated into tangible actions and recommendations. Leveraging the capabilities of the LMS, educational institutions and stakeholders can embark on informed data-driven decision-making. This process is pivotal in enhancing the learning environment and tailoring personalized support to learners, thereby optimizing the educational experience [10].

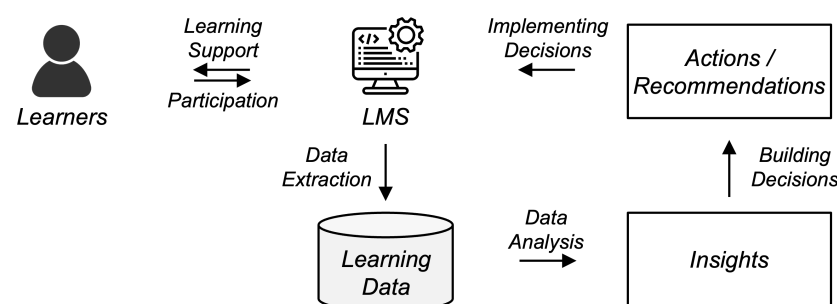


Figure 1. An overview of learning decision support framework.

Regarding this framework, it is of significant importance to derive insights from learning data. Hence, there have been efforts to deal with that [11–16]. More specifically, research on learning analytics is frequently centered around Massive Open Online Courses (MOOCs), which are online courses designed to accommodate many participants over the Internet. MOOC-based learners are typically characterized by self-regulated learning and a degree of freedom from strict oversight. However, this approach is associated with challenges such as lower academic performance and higher dropout rates. Based on these characteristics, research has explored areas such as identifying at-risk students [13], developing intervention systems to enhance learner engagement [14], and predicting student performance using machine-learning models [15,16]. However, the central emphasis of these endeavors has predominantly been on learning outcomes—constructing predictive models aimed at discerning the salient features that influence these outcomes [15,16].

In addition, a notable research gap exists in understanding learning behavior, particularly in distinguishing between self-regulated and guided learning based on self-efficacy. Within this domain, self-regulated learning is characterized by various components such as self-planning, self-monitoring, and self-assessment [17]. It is imperative to note that the practice of self-monitoring significantly influences learning behaviors [18]. To this end, scholarly efforts, as exemplified by [19] in analyzing student quiz-taking behaviors within Learning Management Systems, and categorization of learners based on engagement and dropout rates in [20], have begun to shed light on these behaviors. Additionally, investigation into the correlation between self-regulated learning behaviors and academic outcomes in [20] underscores the importance of understanding these dynamics. Nonetheless, a comprehensive grasp of the intricate learning patterns and their implications for academic performance remains a pressing scholarly necessity.

To deal with these research gaps, we formulate three research questions: (i) What are the behavioral patterns that characterize learner engagement and activity within the learning environment? (ii) Which elements, such as learner demographics, behavioral traits, or content-related factors, exert an influence on these learning patterns? (iii) How do

these learning patterns correlate with learner performance, and what implications do these relationships have for educational outcomes?

This paper endeavors to address the aforementioned research questions through a case study of an empirical analysis of learner interactions within a learning system in the context of English as a Foreign Language (EFL). In EFL learning, innovative approaches have been explored through learning analytics. These include the development of a recommender system designed explicitly for EFL vocabulary acquisition [21], the deployment of a mind-map guided AI chatbot system aimed at enhancing learners' speaking abilities [22], and the examination of variables influencing the oral performance of EFL students [23]. Such studies represent significant progress in online EFL education, demonstrating the potential of technology-enhanced learning interventions. However, despite these advancements, there is a notable research gap in the detailed analysis of EFL learning behaviors and patterns through the examination of log data.

We employ a multifaceted approach that integrates various analytical techniques, including machine learning, process mining, and statistical methodologies. Notably, our analytical toolkit encompasses hierarchical clustering [24] for pattern identification, process discovery for uncovering interaction models [25–27], and the Mann–Kendall test [28,29] for assessing statistical trends and relationships with learning outcomes.

The rest of this paper is structured as follows. In Section 2, we introduce the data used in this case study and the relevant methods. Section 3 delivers the results of the application of our method. Section 4 discusses the significance and the limitations of our approach. Finally, Section 5 concludes the paper by summarizing our work.

2. Materials and Methods

This section outlines our approach to identifying learning patterns from English as a Foreign Language (EFL) service data. Section 2.1 covers the collection and standard format of the learning log data. In Section 2.2, we introduce the framework of our approach. The following sections describe the methods in detail: Section 2.3 focuses on identifying distinct learning patterns, Section 2.4 explores these patterns and their characteristics, and Section 2.5 evaluates the impact of these patterns on learning outcomes.

2.1. Materials

In this study, we utilize data collected from the ReadingN service provided by iPortfolio Inc., Seoul, South Korea. Figure 2 provides a snapshot of this service. This service targets preschool and elementary school children, offering a structured learning approach with globally recognized best-selling English books for children.

The curriculum offered by this service is illustrated in Figure 3, showcasing a collection of renowned book series such as the Oxford Reading Tree (ORT), Big Cat, Oxford Readers Collection (ORC), and Bob Books. Each series encompasses a range of books distinguished by their difficulty levels. It is important to note that learners are not required to start with the most basic levels; they can choose books based on their proficiency. However, each book is structured around a sequence of stages: Warm Up, Listen Up, Read, Speak Up, and Wrap Up. Learners are encouraged to adhere to this structured approach to maximize their learning experience. More specifically, “Warm Up” introduces essential vocabulary with images, “Listen Up” combines pictures and audio for context without text, and “Read” presents the text with native speaker audio and a feature for pronunciation practice. The “Speak Up” focuses on speaking and testing pronunciation accuracy with critical sentences, producing the pronunciation scores, while “Wrap Up” tests comprehension with quizzes. This flexible structure allows learners to move sequentially, focus on specific stages, or review as needed.

Then, we introduce the learning event logs recorded and accumulated within learning management systems. These logs are pivotal for our analysis, providing granular details about learner interactions and activities. Table 1 illustrates a fragment of a learning event log.



Figure 2. Screenshot of Stage 3 ('Read') in the ReadingN Service.

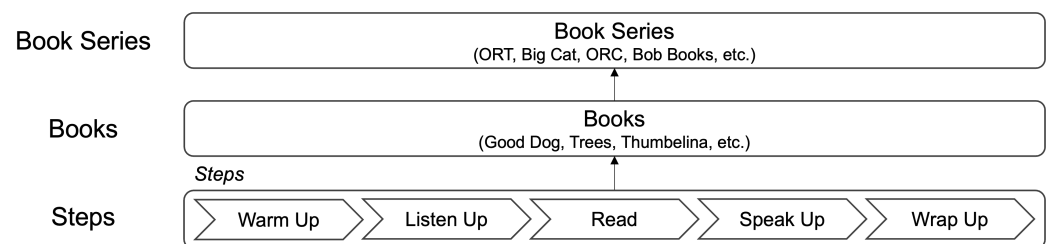


Figure 3. The overall curriculum of this service.

Table 1. A fragment of a learning log.

Event ID	Learner ID	Case ID	Activity	Timestamp	Book ID	Scores
e1	l1	c1	Warm Up	1 January 2023 09:05:00	b1	-
e2	l1	c1	Listen Up	1 January 2023 12:13:00	b1	-
e3	l1	c1	Read	8 January 2023 15:20:00	b1	-
e4	l1	c1	Speak Up	8 January 2023 17:31:00	b1	0.7046
e5	l1	c1	Wrap Up	9 January 2023 08:17:00	b1	-
e6	l2	c2	Warm Up	15 January 2023 13:42:00	b1	-
e7	l2	c2	Listen Up	22 January 2023 11:16:00	b1	-
e8	l2	c2	Speak Up	22 January 2023 13:22:00	b1	0.5607
e9	l1	c3	Warm Up	29 January 2023 18:16:00	b2	-
e10	l1	c3	Warm Up	29 January 2023 18:35:00	b3	-
e11	l1	c3	Listen Up	29 January 2023 18:49:00	b3	-
e12	l1	c3	Read	29 January 2023 18:54:00	b3	-
e13	l1	c3	Speak Up	29 January 2023 19:17:00	b3	0.8275
e14	l3	c4	Warm Up	3 February 2023 10:20:00	b4	-
e15	l3	c4	Listen Up	3 February 2023 10:24:00	b4	-
...	...	...	...	...	...	...

This event log comprises a record of learning events that are the individual actions performed by learners within the LMS. Each event is associated with individual cases representing the learners. In addition, the log includes essential details such as the specific activity undertaken, the corresponding timestamp, and activity-relevant information, including the materials used, such as books and series. Additionally, when available, the performance metrics associated with each activity are incorporated into the log, enriching the dataset with insights into learners' achievements and progress. Based on these, we define a general data structure for event logs as formulated in Definition 1.

Definition 1 (Learning Event Log). Let U_{ev} be the universe of event identifiers, U_{ca} be the universe of case identifiers, U_{act} be the universe of activities, U_{time} be the universe of timestamps, $U_{learner}$ be the universe of learners, U_{att} be the universe of attributes, and U_{va} be the universe

of attribute values. A learning event log is a tuple $LEL = (E, C, Att, \pi)$, where $E \subseteq U_{ev}$ is a set of events, $C \subseteq U_{ca}$ is a set of cases, $Att \subseteq U_{att}$ is a set of attributes, and $\pi \in E \rightarrow (Att \rightarrow U_{va}) \cup C \rightarrow (Att \rightarrow U_{va})$ is a mapping function. For attributes $x \in \text{dom}(\pi(e))$ and $y \in \text{dom}(\pi(c))$, $\pi_x(e) = \pi(x)(e)$ and $\pi_y(c) = \pi(y)(c)$ are the attribute values of x and y for event $e \in E$ and case $c \in C$, respectively. Any learning event log LEL has four mandatory attributes, case, act, time, trace, and learner $\subseteq U_{att}$, such that, for any event e and case c :

- $\pi_{case}(e) \in C$ is the case that e belongs to,
- $\pi_{act}(e) \in U_{act}$ is the activity e refers to,
- $\pi_{time}(e) \in U_{time}$ is the time at which e occurred,
- $\pi_{trace}(c) \in E^*$ is the trace producing a partial sequence of events for c , and
- $\pi_{learner}(c) \in U_{learner}$ is the learner refers to the person associated with c .

Within the learning event log, several supplementary event attributes are provided, as illustrated in Table 1. These extra characteristics, such as Book ID, Series ID, and Scores, are linked to the recorded events and offer additional layers of information.

2.2. Overview

We enumerate an overview of our framework, which comprises three key phases: pattern identification, exploration, and evaluation. Figure 4 visually represents this overview. Pattern identification (Phase 1) aims to identify distinct patterns of learning behaviors recorded in the learning data. To achieve this, we utilize the learning event log, formalized in Definition 1, and apply a trace-based encoding technique. Subsequently, we employ hierarchical clustering with Levenshtein distance to group the encoded learning behaviors. We then extract meaningful patterns guided by expert-defined rules. Furthermore, the process model discovery technique is applied to discover process models for learning patterns. Pattern exploration (Phase 2) aims to explore the derived learning patterns. We first perform the single-view exploratory analysis to investigate the relationships between the learning patterns and other characteristics, e.g., learners or course materials. Then, the multi-view exploratory analysis is conducted to understand which attributes are more related to learning patterns, followed by a post-hoc analysis. The final step, pattern evaluation (Phase 3), centers on assessing the identified patterns' impact on learning achievements. To address this, we initially analyze the performance trends using the Mann–Kendall test. Subsequently, we conduct a comparative analysis between the performance trends and the identified learning patterns.

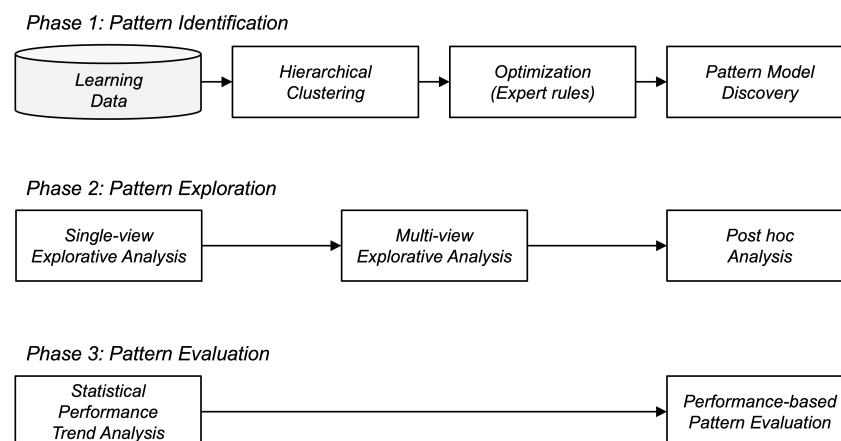


Figure 4. An overview of the proposed framework.

2.3. Pattern Identification

This phase is dedicated to constructing learning patterns derived from the learning behaviors of cases, utilizing the learning event log. Here, we define learning patterns as the

flows of learning steps. Therefore, it focuses on the traces of cases, allowing us to preserve the full spectrum of behavioral information without any loss.

In our method, these traces are used directly as inputs for clustering, which is instrumental in grouping similar learning behaviors. To determine the similarity between two sequences of learning behaviors, we utilize the Levenshtein distance, commonly referred to as the edit distance [30]. This distance metric is particularly effective for measuring how closely related two sequences are, offering a detailed perspective on the nuances of learning behaviors.

The Levenshtein distance is designed to quantify the similarity between two strings. Specifically, when two words are identical, their Levenshtein distance is 0, while if there are substantial differences, the distance approaches 1. In our context, each activity within a trace is treated as a single character in a string. Thus, two nearly identical activity sequences will have a Levenshtein distance approaching 0. This approach is encapsulated in the following distance formula between two traces, $\pi_{trace}(c_a)$, $\pi_{trace}(c_b)$ of cases $c_a, c_b \in C$:

$$Lev[t(c_a), t(c_b)] = \begin{cases} 0 & \text{if } t(c_a) = t(c_b) \\ Lev[t(c_a)_{1:|t(c_a)|-1}, t(c_b)_{1:|t(c_b)|-1}] & \text{if } t(c_a)_{|t(c_a)|} = t(c_b)_{|t(c_b)|} \\ 1 + \min \begin{cases} Lev[t(c_a)_{1:|t(c_a)|-1}, t(c_b)] \\ Lev[t(c_a), t(c_b)_{1:|t(c_b)|-1}] \\ Lev[t(c_a)_{1:|t(c_a)|-1}, t(c_b)_{1:|t(c_b)|-1}] \end{cases} & \text{otherwise} \end{cases} \quad (1)$$

where $t(c_a)$ and $t(c_b)$ represent the traces $\pi_{trace}(c_a)$ and $\pi_{trace}(c_b)$ of cases c_a and c_b , $|t(c_a)|$ and $|t(c_b)|$ represent the lengths of the traces $t(c_a)$ and $t(c_b)$, and $t(c_b)_{1:|t(c_b)|-1}$ and $t(c_a)_{1:|t(c_a)|-1}$ represent the subsequences of traces $t(c_a)$ and $t(c_b)$ from the first activity to the second-to-last activity.

Based on the formula, we can construct a distance matrix, D , among all cases as follows:

$$D = \forall_{c_i \in C} \forall_{c_j \in C} \{ Lev[\pi_{trace}(c_i), \pi_{trace}(c_j)] \} \quad (2)$$

where D is a $|C| \times |C|$ matrix to reflect the distances between each pair of cases, computed using the Levenshtein distance formula.

We perform hierarchical clustering using the constructed distance matrix to create groups exhibiting similar learning behaviors. In the specific context of our research, the process begins by assigning each case to its cluster, thereby forming ($|C|$) initial clusters, where each cluster contains only one object. The hierarchical clustering then follows these iterative steps until only a single cluster remains:

- (i) identifying the pair of clusters (CL_k, CL_l), i.e., a group of cases, that have the smallest distance ($D[CL_k, CL_l]$) between any of their respective cases, where $D[CL_k, CL_l] = \sum_{c_j \in CL_l} Lev[\pi_{trace}(c_i), \pi_{trace}(c_j)]$ is the average pairwise distance between two clusters,
- (ii) combining the identified two clusters (CL_k, CL_l) into a single cluster,
- (iii) updating the distance matrix (D) to reflect the new distances between the newly formed cluster and the remaining clusters.

By repeating these steps, the hierarchical clustering algorithm gradually amalgamates clusters based on their behavioral similarities until converging into a comprehensive cluster encompassing all cases.

In addition, selecting an appropriate linkage method is crucial in the updating step of the hierarchical clustering process. Linkage methods determine how the distance between clusters is measured and include single, complete, and average linkage options. For our study, we opt for the average linkage method, and the corresponding objective function for our clustering method can be formulated as follows:

$$\min \left\{ \frac{1}{|CL|(|CL|-1)} \sum_{k=1}^{|CL|} \sum_{l \neq k} D[CL_k, CL_l] \right\} \quad (3)$$

where $|CL|$ represents the number of clusters and $D[CL_k, CL_l]$ is the average pairwise distance between two clusters, CL_k and CL_l .

As a direct outcome of this process, we can identify key shared learning patterns within the data. However, we can further refine and concretize these learning patterns in this context by incorporating expert rules. This approach integrates domain-specific knowledge and expertise into the analysis, allowing for a more nuanced and contextually relevant interpretation of the patterns. Based on these, we can define a function, $learnp \in C \rightarrow LP$, where C is a set of cases and LP is a set of learning patterns. Therefore, it maps each case onto the possible learning patterns.

Leveraging the distinct activity sequence groups identified as learning patterns, we elucidate the overall flow within each cluster by discovering a process model. For this purpose, we utilize process mining, specifically focusing on its process discovery functionality. Process mining is a scientific discipline that derives process-related insights from event logs and encompasses three key functionalities: process discovery, conformance checking, and enhancement. Among them, process discovery is particularly relevant to our study as it facilitates the automatic generation of process models directly from data, minimizing the need for manual input.

In this context, a wide range of process discovery algorithms is available, each with its corresponding process model notation. In our research, we employ the Direct Follows Model (DFM) [31], denoted as $DFM = (N, E)$, where $N = \Sigma \cup \{\text{start}, \text{end}\}$ represents the set of nodes and $E = N \times N$ is the set of edges, and the DFM discovery algorithm. The choice of the DFM and its discovery algorithm is driven by its significant strengths: simplicity and ease of understanding.

The DFM is particularly adept at providing clear and concise visualizations of processes, making it an excellent tool for representing learning patterns in an easily interpretable manner. This approach allows us to effectively communicate the nature and structure of the learning behaviors within each cluster, offering valuable insights into the learning process as it unfolds within the data.

2.4. Pattern Exploration

This section aims to explore and understand the connections between learning patterns and various elements, explicitly focusing on the relationship between these patterns and learners or course materials. Note that we limit course materials to books. To achieve this, we implement the single-view and multi-view exploratory analysis.

In the context of the single-view analysis, we employ cross-tabulation [32] and chi-squared tests [33] to analyze how learning patterns are distributed across learners and books. The cross-tabulation is a technique for summarizing categorical data to understand the relationships between variables. For instance, to examine the relationship between learning patterns and learners or books, we construct contingency tables for learners ($CT_{learners}$) and books (CT_{books}). These tables are formulated as follows:

$$CT_{learners} = \forall_{l_i \in LN} \forall_{lp_j \in LP} \left\{ \sum_{c \in C} 1 | \pi_{learner}(c) = l_i \wedge learnp(c) = lp_j \right\} \quad (4)$$

$$CT_{books} = \forall_{b_i \in B} \forall_{lp_j \in LP} \left\{ \sum_{c \in C} 1 | \pi_{book}(c) = b_i \wedge learnp(c) = lp_j \right\} \quad (5)$$

where $LN \subset U_{learner}$ is the set of learners, where $B \subset U_{val}$ is the set of books; LP is the set of learning patterns; and $\pi_{learner}(c)$, $\pi_{book}(c)$, and $learnp(c)$ are the functions that map a case to a learner, a book, and a learning pattern, respectively.

Next, the chi-squared test is utilized by calculating the chi-square statistic (χ^2) for each table in the following manner:

$$\sum_{i=1}^{|V|} \sum_{j=1}^{|LP|} \frac{(CT_{ij} - E_{ij})^2}{E_{ij}} \quad (6)$$

where $|V|$ is the number of variables, i.e., learners or books, $|LP|$ is the number of learning patterns, CT_{ij} is the observed frequency of the i th variable and j th learning pattern from the contingency tables, including $CT_{learners}$ and CT_{books} , and E_{ij} is the expected frequency of the i th variable and j th learning pattern, computed by $\frac{\sum_{i=1}^{|V|} CT_{ij} \times \sum_{j=1}^{|LP|} CT_{ij}}{\sum_{i=1}^{|V|} \sum_{j=1}^{|LP|} CT_{ij}}$. Then, we make a statistical determination by comparing the critical value, typically at a significance level such as 0.05, with the degrees of freedom (df) computed as $(|V| - 1) \times (|LP| - 1)$. This assessment allows us to ascertain whether learners or behaviors exhibit even distributions across learning patterns.

For the multi-view analysis, we examine how the identified learning patterns correlate with learners or course materials. In this context, the diversity indices for multiple classes are utilized, as illustrated in Table 2. Then, we can build the set of diversity values for learners ($D_{learner}$) and books (D_{book}). Expressly, a higher value associated with a learner or book signifies a broader spectrum of learning patterns linked to them, whereas a lower value suggests a concentration in fewer learning patterns.

Table 2. The diversity indices with multiple classes.

Diversity Indices	Formula (div)
Hill numbers [34]	$\left(\sum_{j=1}^{ LP } p_j^q \right)^{1/(1-q)}$
Shannon index [35]	$-\sum_{j=1}^{ LP } p_j \ln(p_j)$
Gini–Simpson index [36]	$1 - \sum_{j=1}^{ LP } (p_j)^2$
Berger–Parker index [37]	$\frac{N_{max}}{N}$

Subsequently, we apply the t -test to determine if there is a significant difference between the diversity values of learners and books. The test statistic is calculated as:

$$t = \frac{E(D_{learner}) - E(D_{book})}{\sqrt{\frac{var(D_{learner})}{|LN|} + \frac{var(D_{book})}{|B|}}} \quad (7)$$

where $E(D_{learner})$ and $E(D_{book})$ are the sample means of two diversity values, $var(D_{learner})$ and $var(D_{book})$ are the sample variances of two diversity values, and $|LN|$ and $|B|$ are the number of learners and books. Then, the calculated t -value is compared against the critical value obtained from the t -distribution with the degrees of freedom, computed as $(\frac{var(D_{learner})}{|LN|} + \frac{var(D_{book})}{|B|})^2 / (\frac{var(D_{learner})}{|LN|-1} + \frac{var(D_{book})}{|B|-1})$. Note that we can also consider the Mann–Whitney U Test [38] as an alternative method, depending on the normality of the distribution. In addition, if we have more attributes we can consider the one-way analysis of variance (ANOVA) [39] or Kruskal–Wallis H test [40].

The results obtained from this analysis enable us to discern the overarching characteristics of learning patterns. Moreover, we can delve deeper into the analysis if we can access more granular attributes related to learners and course materials as the post-hoc analysis.

This deeper analysis could involve exploring the intricate relationships between learning patterns and these detailed features or constructing predictive models for learning patterns using these additional features as inputs.

2.5. Pattern Evaluation

In this last phase, we aim to establish a connection between learning patterns and academic achievements. In this context, academic achievements can be determined in various formats based on course characteristics or learning systems. In this research, our materials quantify academic achievements as pronunciation scores obtained during evaluation activities. Note that our research context presents a unique aspect that each learner contributes multiple records related to course materials. Therefore, to link learning patterns to academic scores, we frame the problem as how a sequence of learning patterns exhibited by users influences the change in their academic scores. This user-centric analysis is compelled by the fact that each user possesses a distinct initial level of speaking skills.

To address this, our initial step is to gain insights into the trends within performance scores, which may involve increases, decreases, or remaining relatively stable. To achieve this, we employ the Mann–Kendall statistical test [28,29]. This method is designed to assess whether there exists a monotonic trend, or lack thereof, within the performance scores.

Given that a series of learning patterns, $\mathbf{lp}_i = (lp_{i1}, lp_{i2}, \dots, lp_{in})$, and the corresponding scores, $\mathbf{s}_i = (s_{i1}, s_{i2}, \dots, s_{in}) = (\pi_{score}(c_{i1}), \pi_{score}(c_{i2}), \dots, \pi_{score}(c_{in}))$, associated with the cases, $c_{i1}, c_{i2}, \dots, c_{in}$, engaged by a learner, l_i , the Mann–Kendall statistic for scores, S , is calculated as follows:

$$S = \sum_{j=1}^{n-1} \sum_{k=j+1}^n \text{sgn}(s_{ij} - s_{ik}) \quad (8)$$

where n is the length of cases of learner l_i and sgn is the sign function. Here, it yields values of 1, 0, and -1 when $s_{ij} - s_{ik} > 0$, $s_{ij} - s_{ik} = 0$, and $s_{ij} - s_{ik} < 0$, respectively. Then, we compute the Mann–Kendall test statistic, Z_{MK} , as follows:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & S < 0 \end{cases} \quad (9)$$

where $V(S) = \frac{1}{18} \left(n(n-1)(2n+5) - \sum_{k=1}^p t_k(t_k-1)(2t_k+5) \right)$ is the variance of the Mann–Kendall statistic, S . Here, p is the number of tied groups and t_k is the size of the k -th tied group. By computing Z_{MK} and comparing it with the critical value with a given significance level, we can draw conclusions about the trend in achievement scores. In more detail, we can determine whether the scores exhibit a monotonic trend, an upward trend, or a downward trend.

Then, we can leverage a range of sequence analysis techniques to reveal sequential patterns associated with upward or downward trends. In essence, we can detect a series of learning patterns that affect changes in academic achievements. Here, we can explore methods such as Prefixspan, SPADE (Sequential Pattern Discovery using Equivalence classes), GSP (Generalized Sequential Pattern), etc. [41]. Furthermore, we assess the significance of these sequences based on their support values, aiding in the evaluation of their relevance and importance.

3. Results

In this section, we present our case study results by applying our framework to a specific learning dataset. We begin with dataset details and preprocessing (Section 3.1), followed by the framework's three phases: pattern identification using hierarchical clustering (Section 3.2), pattern exploration with statistical analysis (Section 3.3), and pattern evaluation against learning achievements (Section 3.4). It presents the effectiveness of our framework in analyzing learning patterns with educational data.

3.1. Descriptive Statistics

As elaborated in Section 2.1, we obtained a learning event log from a platform dedicated to English language book learning. This dataset conforms to the established format of a learning event log, as delineated in Definition 1 and Table 1. Notably, the dataset encompasses a defined temporal scope spanning 7 months, starting from 1 January 2023 and concluding at 26 July 2023. During this time-frame, the service was utilized by 182 distinct learners, resulting in the inclusion of 28,192 individual cases, denoting discrete instances of the learning process. Furthermore, the log records a cumulative total of 193,663 learning events, yielding an average of 6.87 events per individual case.

Furthermore, the dataset encompasses case and event attributes, encompassing learners, traces, books, series, and scores. In this context, learners denote the unique identifiers linked to individual cases, whereas traces represent partial sequences of activities. Note that our characterization of learning patterns is rooted in analyzing these traces. Additionally, the dataset contains information pertaining to books and series that are relevant to the educational content accessed by the learners. Lastly, the scores attribute exclusively pertains to the 'Speak Up' activity, signifying the outcomes of pronunciation evaluations for specific sentences within the learning process.

3.2. Results for Pattern Identification

To initially categorize similar learning behaviors, we conducted hierarchical clustering using the Levenshtein distance metric. Furthermore, we implement a threshold value of 2500, which leads to the discovery of 43 clusters exhibiting similar learning behaviors.

Subsequently, through the qualitative evaluation, we identified and derived a set of nine representative learning patterns, as follows:

- Sequential Learning (P1). Learners engage with the learning contents in a linear fashion, progressing through it once. As a result, it strictly adheres to the specific sequence of activities: Warm Up, Listen Up, Read, Speak Up, and Wrap Up.
- Repetitive Sequential Learning (P2). Learners systematically revisit the content in a linear order multiple times. Within this pattern, learners frequently strive to study the same book repeatedly, following a predefined sequence, which may take the form of the following sequence: <Warm Up, Listen Up, Read, Speak Up, Wrap Up, Warm Up, Listen Up, Read, Speak Up, Wrap Up>.
- Sequential Learning with Additional Phases (P3). Learners adhere to a structured linear learning path while integrating additional phases. Specifically, it entails a sequential progression from Warm Up to Wrap Up, with the inclusion of extra activities as follows: <Warm Up, Listen Up, Read, Speak Up, Wrap Up, Speak Up, Wrap Up> or <Warm Up, Warm Up, Listen Up, Read, Speak Up, Wrap Up>.
- Single-Phase Learning (P4). Learners engage in a singular learning activity, and they do so only once. This pattern is distinguished by learners exploring and perusing the content within a book. In this context, it may involve patterns such as <Warm Up> or <Speak Up>.
- Repetitive Single-Phase Learning (P5). Analogous to single-phase learning, learners complete a single phase. The distinguishing feature here is that they revisit this phase multiple times. Within this context, patterns such as <Warm Up, Warm Up, Warm Up> may be observed, signifying the repetition of a specific learning activity.
- Partial-phase Learning (P6). Learners participate in partial phases of learning, yet partial phases are executed a restricted number of times. This pattern represents a typical self-regulated learning approach that does not encompass all steps but selectively involves some of them. However, this pattern is characterized by being undertaken only a few times rather than repetitively, such as <Read, Speak Up> or <Listen Up, Read>.
- Repetitive Partial-phase Learning (P7). Learners involve themselves in and successfully conclude partial phases of learning, and these partial phases are completed

numerous times. Consequently, within this context learners follow sequences such as <Read, Read, Read, Read, Read, Speak Up, Speak Up, Speak Up, Speak Up, Speak Up>.

- Non-sequential Learning (P8): Learners interact with the content in a non-linear or disordered fashion, without adhering to a specific sequence. This learning pattern, akin to single-phase learning, is a form of self-regulated learning. However, it differs in that it encompasses all phases of the learning contents but does not follow a predetermined order. Consequently, it follows patterns such as <Warm Up, Listen Up, Wrap Up, Speak Up, Read>, where the phases are covered, but their sequence is not strictly adhered to.
- Non-sequential Learning with Additional Phases (P9): Learners interact with the content in a disorganized manner, incorporating additional phases into their learning process. This pattern follows the sequence <Warm Up, Listen Up, Wrap Up, Speak Up, Read, Speak Up, Speak Up>.

In addition, we determined the distribution of learning patterns, as illustrated in Table 3. Among the nine identified learning patterns, sequential learning (P1) stands out as the most prevalent, constituting 35%. Furthermore, it is noteworthy that a substantial number of learners opt for sequential learning with additional phases (P3) and repetitive sequential learning patterns (P2). This suggests a strong inclination among individuals to adhere closely to the predefined learning sequences and guidelines.

Table 3. The distribution of the derived learning patterns.

Pattern No.	Description	Frequency	Rate
P1	Sequential Learning	9868	35.0%
P2	Repetitive Sequential Learning	3138	11.1%
P3	Sequential Learning with Additional Phases	5031	17.9%
P4	Single-phase Learning	2726	9.7%
P5	Repetitive Single-phase Learning	1384	4.9%
P6	Partial-phase Learning	2667	9.5%
P7	Repetitive Partial-phase Learning	595	2.1%
P8	Non-sequential Learning	1233	4.4%
P9	Non-sequential Learning with Additional Phases	1550	5.5%

To enhance the comprehensibility and depth of our analysis, we advanced our study by visualizing process models corresponding to each identified learning pattern. Figure 5 depicts these process models for the learning patterns. For this purpose, we employed Fluxicon DISCO [42], a prominent tool in the field of process mining. Here, each node within process models represents a specific activity, while the edges denote the sequential flows between these activities. In addition, the darkness and thickness of nodes and edges indicate the frequency of occurrence; thus, darker nodes and thicker edges imply a higher frequency.

Basically, the process models allow us to discern the nuanced characteristics of each learning pattern. Notably, Figure 5a–c predominantly exhibit sequential flows, reflecting the primary behavior within these learning patterns. Specifically, Figure 5a,b demonstrate learning patterns characterized by sequences with repetitive elements, while Figure 5c expands this to include additional, more complex steps. This complexity is evidenced by recursive flows, indicating a return to earlier stages in the learning process, suggesting a more iterative and dynamic learning behavior.

Additionally, Figure 5d,e delineate learning patterns characterized by single-phase activities, both with and without repetition. These process models reveal that all identified learning flows are confined to a singular phase within the learning process. A notable aspect of these patterns is the prevalence of specific activities such as ‘Read’ and ‘Level Up’.

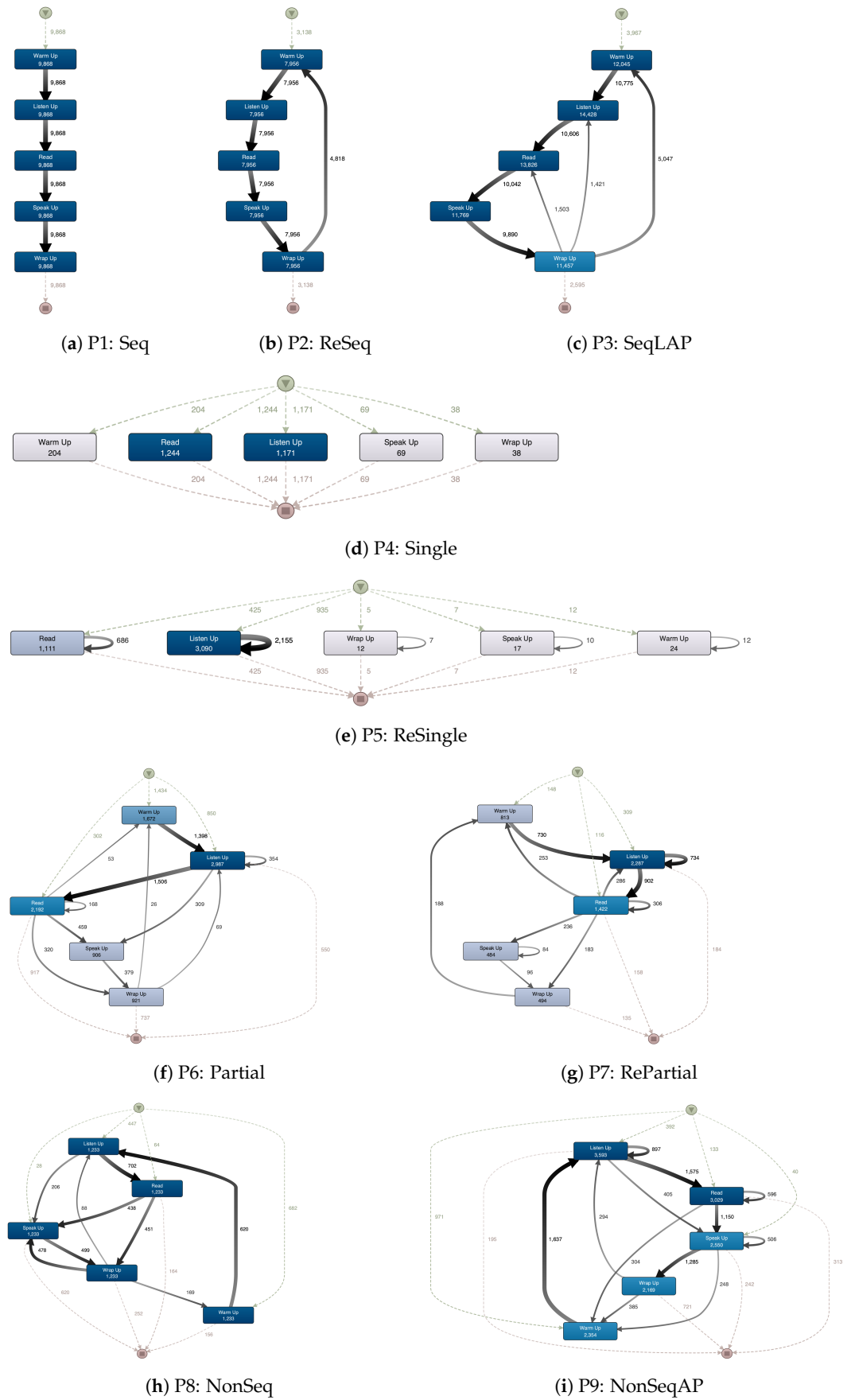


Figure 5. Process models for learning patterns.

Moreover, Figure 5f,g illustrate the learning patterns that we categorize as partial-phase patterns. A notable observation in these figures is the darkening of specific nodes compared to previously discussed sequential learning patterns. In addition, considering the thickness of the edges we can identify that the most common sequences of activities are <Warm Up, Level Up, Read>. This sequence indicates a typical progression pattern within these partial-phase learning patterns.

Finally, we also explored the process models for non-sequential learning patterns, as depicted in Figure 5h,i. These models contrast markedly with the sequential learning patterns, encompassing various learning flows with a greater degree of self-regulated behaviors exhibited by learners.

3.3. Results for Pattern Exploration

In this section, we explain the results derived from analyzing the correlation between learning patterns and the attributes of learners as well as the utilized instructional materials, i.e., books. Initially, we employed the chi-squared goodness of fit test to ascertain the uniformity in the distribution of learning patterns among various learners and book types. This necessitated the construction of contingency tables for both learners and books through cross-tabulation, which facilitated a structured comparison of learning patterns across different categories.

Subsequently, we applied the chi-square test, framed within two hypotheses. The null hypothesis (H_0) posited that there is no preferential association of learning patterns with specific learners or books. Conversely, the alternative hypothesis (H_a) contended that learning patterns are distinctly prevalent among particular learners or books. It yielded chi-squared statistics of 53,161.79 for learners and 21,838.59 for books, surpassing the critical values, 1537.63 and 17,683.76, at a 0.05 significance level; in other words, the p -values for both cases are less than 0.05.

The statistical significance of these results led to the rejection of the null hypothesis, thereby underscoring a non-random distribution of learning patterns. This implies that learning patterns are influenced by factors beyond mere learner characteristics or book types, suggesting a complex interplay of various educational elements in shaping learning behavior.

After that, we utilized multiple diversity indices across various cases to gain a deeper understanding of the association between learning patterns and either learners or books. These indices included the Hill number with q of 0, Shannon, Gini–Simpson, and Berger–Parker indices. For each learner and book, we calculated diversity scores based on these indices. The aggregated diversity values for both learners and books are presented in Table 4.

Table 4. The comparison of diversity indices for learners and books.

Diversity Indices	Learners			Books		
	Average	Minimum	Maximum	Average	Minimum	Maximum
Hill	6.302	1.000	9.000	3.786	1.000	9.000
Shannon	1.686	0.000	2.689	1.328	0.000	3.031
Gini–Simpson	0.570	0.000	0.826	0.473	0.000	0.871
Berger–Parker	0.566	0.236	1.000	0.601	0.161	1.000

To assess whether learning patterns are more associated with either learners or books, we conducted a t -test. The null hypothesis for this test posited that the diversity values for learners and books are equivalent. In contrast, the alternative hypothesis suggested that these values differ significantly. The outcomes of the Student's t -test are detailed in Table 5.

The results showed that the t -statistic values for most indices exceeded the critical values at a significance level of 0.05, indicating a significant difference in diversity values between learners and books. Specifically, lower values in the Shannon index, Gini–Simpson index, and Hill numbers imply less diversity, whereas a higher value in the Berger–Parker

index indicates reduced diversity. In all instances, the diversity of learning patterns associated with books was lower than that of learners. This leads us to conclude that learning patterns are more closely associated with books rather than with individual learners, suggesting a more decisive influence of the instructional material on learning patterns.

Table 5. The *t*-test results for diversity indices between learners and books.

	Hill	Shannon	Gini–Simpson	Berger–Parker
<i>t</i> -statistic	17.676	4.821	4.049	1.697
<i>p</i> -value	<0.001	<0.001	<0.001	0.090

Regarding the level of books, for a more refined analysis we grouped these into three categories: low (levels 1–4), moderate (levels 5–8), and high (levels 9–12). Our focus was to examine these categories in the context of learning patterns. Figure 6 in our study illustrates the distribution of learning patterns across these reading levels. In the figure, blue, red, and green bars represent low, moderate, and high levels, respectively, while grey bars depict the overall distribution, as delineated in Table 3.

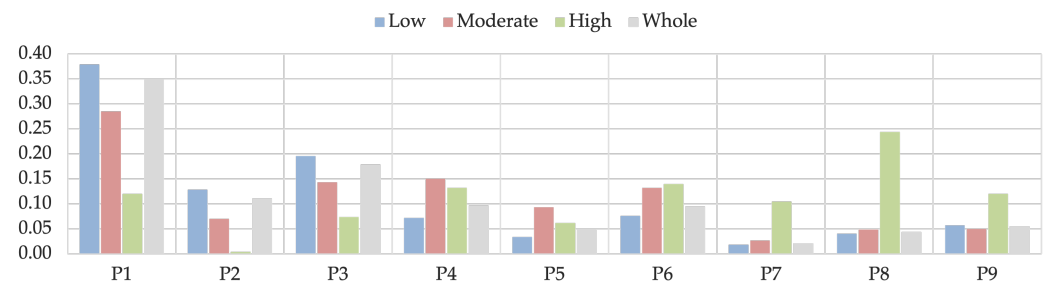


Figure 6. The distribution of learning patterns across the reading levels of books.

The results of our analysis reveal distinct correlations between the level of books and the learning patterns learners adopt. Specifically, for lower-level books, learners tend to engage more with sequential learning patterns P1, P2, and P3. This indicates that with more straightforward material they follow a step-by-step approach. On the other hand, for more challenging books classified as moderate and high levels, learners tend to shift towards less linear learning strategies, as seen in patterns P4 to P7. This shows that they adapt their learning methods to more varied and fragmented approaches as the material gets more complicated. Moreover, learning patterns P8 and P9 suggest that learners adopt self-regulated learning behaviors for the most complex books. This indicates a move towards more independent and flexible learning strategies in response to higher difficulty levels.

Subsequently, we explored the connection between learning patterns and specific book series. Our case study focused on the top four series in our dataset: Oxford Reading Tree (ORT), Big Cat, Oxford Readers Collection (ORC), Bob Books, and a category labeled 'others'. Figure 7 illustrates how different learning patterns are distributed across these series.

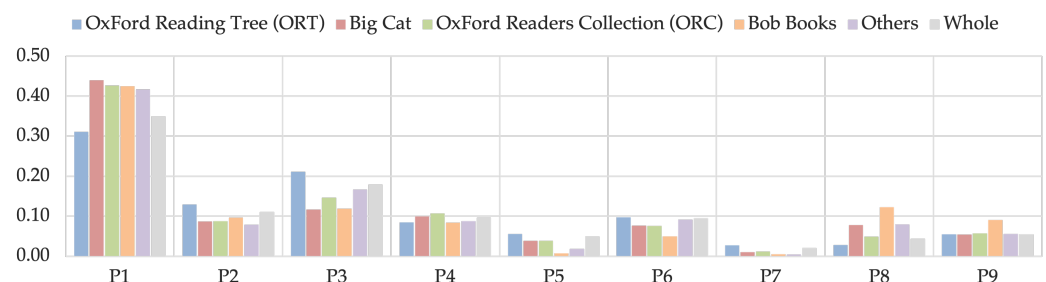


Figure 7. The distribution of learning patterns across the series of books.

From this result, we noted that ORT books tend to have lower engagement in pattern P1 but higher instances of patterns P2 and P3 compared to other series. This suggests that learners using ORT books engage in sequential learning with a tendency towards repetition. In contrast, for patterns P8 and P9, we observed Big Cat and Bob Books show higher values than other series. This indicates that learners using these series exhibit more self-regulated learning behaviors.

3.4. Results for Pattern Evaluation

This section evaluates the identified learning patterns in relation to learning outcomes. For this purpose, we utilized the Mann–Kendall statistical test within the context of our learning log data to assess the learning outcomes of learners. Note that we normalized the pronunciation scores by adjusting them based on the average and standard deviation of scores for each specific book, assuming the pronunciation scores vary depending on the books used. This normalization process ensures that the scores are comparable across different books, providing a more accurate and fair assessment of learners' performance. After that, we proceeded with the Mann–Kendall statistical analysis to analyze whether the learners' performances were increased, neutral, or decreased.

Our analysis revealed that, out of the total learners studied, 49 (27%) exhibited an increasing trend in their learning outcomes, 43 (24%) showed a decreasing trend, and 90 (49%) maintained a neutral trend. The results of the Mann–Kendall test, particularly for five representative learners with the highest absolute values of trend slopes in each category, are presented in Table 6. Notably, while the slope values for both increasing and decreasing trends were not exceptionally large, the p -values indicated that these trends were statistically significant. In contrast, learners with neutral trends displayed almost zero slope values and high p -values, suggesting no significant change in their learning outcomes over time.

Table 6. Results of the Mann–Kendall analysis for top 5 users.

Trend	User	Slope	Z_{MK}	p -Value
Increase	UI1	0.0131	2.557	0.011
	UI2	0.0076	4.736	<0.001
	UI3	0.0065	4.774	<0.001
	UI4	0.0060	3.825	<0.001
	UI5	0.0042	2.871	<0.001
Decrease	UD1	−0.0086	−3.083	<0.001
	UD2	−0.0053	−3.777	<0.001
	UD3	−0.0037	−6.015	<0.001
	UD4	−0.0035	−3.422	<0.001
	UD5	−0.0034	−2.779	<0.001
Neutral	UN1	≈0.0000	0.062	0.951
	UN2	≈0.0000	−0.012	0.991
	UN3	≈0.0000	0.244	0.807
	UN4	≈0.0000	−0.376	0.707
	UN5	≈0.0000	−0.943	0.346

In addition, Figure 8 provides a graphical representation of these trend changes over time, focusing on the top five subjects as determined by the Mann–Kendall test. Each line in the figure represents a different subject, illustrating the extent and direction of their learning performance trends during the observation period. This visual representation allows for a straightforward comprehension of the changes in scores, highlighting the progression or regression in learning outcomes for each subject.

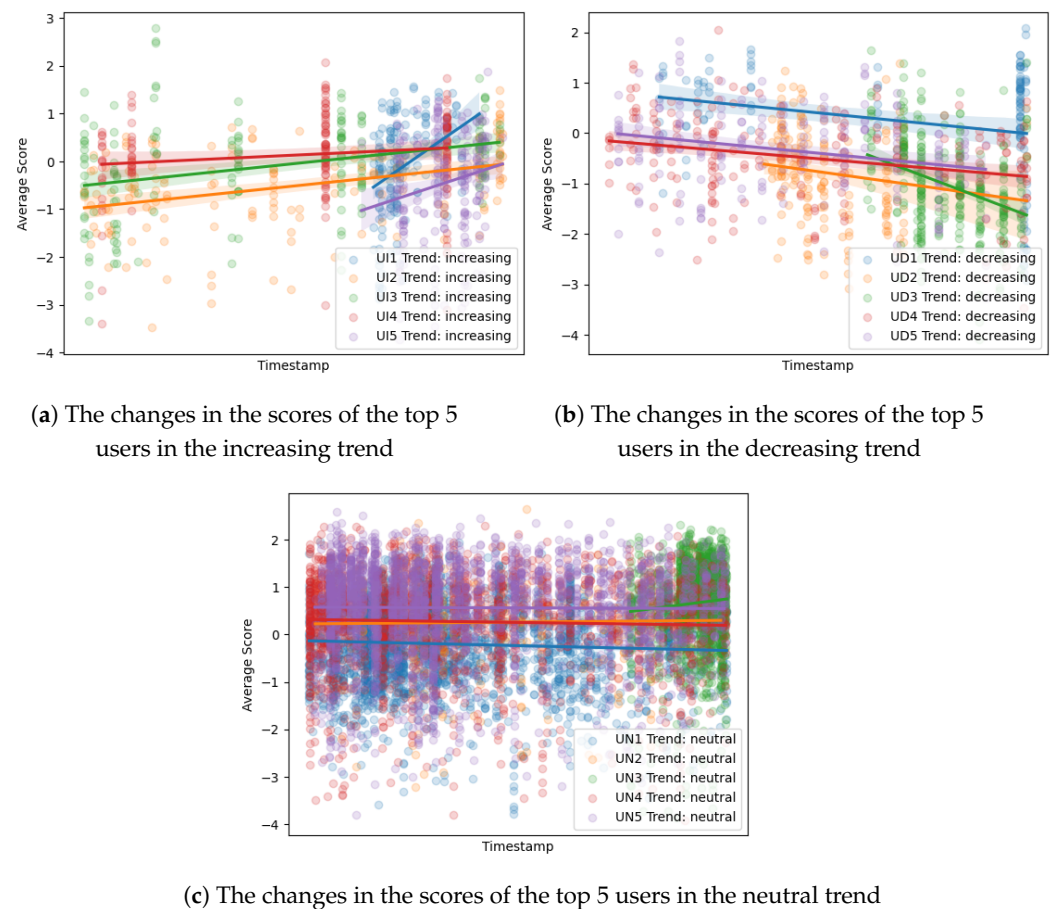


Figure 8. The changes in the scores of users in each group based on the Mann–Kendall analysis.

To further understand the impact of learning patterns on academic performance, we categorized learning outcomes into two groups: ‘increasing’ and ‘neutral or decreasing’. This classification is based on the premise that the direction of change in scores over time—whether they increase or not—is a critical indicator of academic performance. Specifically, an increase in scores indicates improvement, while neutral or decreasing scores are equated with negative or non-improved academic performance.

We then applied sequential pattern mining to our dataset, using the Generalized Sequential Pattern (GSP) method to compare the most common sequences of learning patterns between two groups. Table 7 presents the top 20 sequences of patterns identified for each group.

Overall, the patterns discovered in both groups were broadly similar. A notable observation was that the sequences consisting predominantly of patterns P1 and P3 formed a significant portion of the results. This prevalence is attributed to the non-uniform distribution of learning patterns across the dataset, with specific patterns, e.g., P1 and P3, occurring more frequently than others.

A key finding of interest was the high consistency of pattern P1, exemplified by sequences such as $P1 \rightarrow P1 \rightarrow P1 \rightarrow P1 \rightarrow P1$ and $P1 \rightarrow P1 \rightarrow P1 \rightarrow P1 \rightarrow P1$, which were exclusively present in the non-improved group. This suggests a strong tendency among this group to persist with a singular learning pattern rather than transitioning to other patterns. In addition, in the improved group there was a noticeable trend of transitioning from sequential learning patterns to more self-regulated ones, as indicated by sequences such as $P1 \rightarrow P9$ or $P3 \rightarrow P9$. This implies a dynamic shift in learning strategies within this group, moving from more structured to autonomous learning approaches. On the other hand, the non-improved group displayed sequences that suggested a reverse trend, i.e., $P9 \rightarrow P1$, moving from self-regulated to more sequential patterns. These findings indicate that transition to self-regulated learning patterns may be associated with improved learning outcomes.

Table 7. Top 20 pattern flows for two groups.

	Increasing		Neutral & Decreasing	
	Pattern Sequence	Support	Pattern Sequence	Support
1	P1	0.92	P1	0.88
2	P3	0.92	P3	0.88
3	P1→P1	0.88	P1→P1	0.79
4	P9	0.82	P9	0.76
5	P3→P1	0.76	P3→P1	0.75
6	P3→P3	0.76	P1→P1→P1	0.74
7	P1→P3	0.73	P1→P3	0.72
8	P3→P1→P1	0.71	P1→P1→P1→P1	0.67
9	P1→P1→P1	0.69	P3→P3	0.65
10	P2	0.65	P3→P1→P1	0.64
11	P1→P1→P1→P1	0.63	P1→P1→P3	0.62
12	P6	0.61	P2	0.60
13	P1→P1→P3	0.61	P6	0.60
14	P3→P3→P3	0.61	P1→P3→P1	0.60
15	P1→P1→P1→P3	0.61	P1→P1→P1→P1→P1	0.57
16	P8	0.59	P8	0.53
17	P2→P3	0.57	P3→P1→P1→P1	0.53
18	P3→P9	0.57	P1→P1→P1→P3	0.51
19	P1→P3→P1	0.57	P1→P1→P1→P1→P1→P1	0.51
20	P1→P9	0.55	P9→P1	0.50

In addition, we examined the distribution of learning patterns within these outcome groups. This involved analyzing the proportion of each learning pattern among learners in the improved and non-improved groups. The findings of this analysis are detailed in Table 8.

Table 8. The distribution of learning patterns for two groups.

	P1	P2	P3	P4	P5	P6	P7	P8	P9
Inc.	0.387	0.197	0.228	0.001	0.000	0.002	0.001	0.098	0.063
Neu. & Dec.	0.450	0.145	0.233	0.002	0.000	0.041	0.015	0.043	0.070

The results highlighted noticeable differences in patterns P1, P2, and P8 between the two groups. For patterns P1 and P2, we observed that the ‘improved’ group had a higher engagement in pattern P2 (repetitive sequential learning) but lower engagement in pattern P1 compared to the ‘non-improved’ group. This suggests that learners in the improved group tend to prefer repetitive and iterative learning approaches. Furthermore, in pattern P8 the improved group also demonstrated a higher engagement, indicating a preference for self-regulated learning strategies. To sum up, learners who show improvement are more likely to engage in learning patterns that involve repetition and self-regulation.

4. Discussion

This section offers a comprehensive discussion of our framework and the results from our study. We start by outlining the key findings of our research questions, detailed in Section 4.1. Then, we explore the broader implications of our findings and the limitations inherent in our approach in Section 4.2.

4.1. Discussion for Results

As outlined in Section 1, our study was guided by three research questions. In this section, we provide answers to each of these questions based on the findings from our case study.

In response to the initial research question, our empirical investigation has delineated nine learning patterns. Learners' fidelity determines these learning patterns to instructional guidelines, their propensity for iterative engagement with content, and their selective emphasis on discrete segments of the educational trajectory. Subsequent process discovery [25] facilitated the explication of the distinct characteristics inherent to each learning pattern, which span a spectrum from rigid conformity to prescribed educational sequences to a more autonomous and adaptive learning stance. Such heterogeneity in learning modalities not only encapsulates the multifaceted dimensions of the learning process [43,44] but also substantiates the efficacy of our investigative framework in encapsulating the nuanced dynamics of learner–material interaction [45].

In investigating the second research question, our study aimed to ascertain the variables influencing learning patterns. The analysis revealed a significant revelation: the selection of learning materials, particularly the complexity and series of books, played a more consequential role in shaping learning patterns than the learners' characteristics. This suggests that the qualitative features of educational materials are central to the engagement strategies adopted by learners [46]. Moreover, our study uncovered that the level of difficulty and book series affect learning behaviors. These findings offer critical insights for educators in the selection of learning materials, advocating for a tailored approach to accommodate the diverse requirements of learners [47].

Our research addressed the final question by looking at the link between the types of learning behaviors and students' academic performance. Our findings challenge the idea that following a set learning path guarantees better academic results [48]. In fact, sticking only to such a path, which we called pattern P1, did not always mean students did better. On the other hand, learners who managed their learning process, particularly those in patterns P8 and P9, tended to improve their performance more noticeably. These results are important for educators because they show that they might do better when learners are more in control of their learning [48]. Understanding this can help teachers and curriculum designers create learning experiences that are more flexible and suited to individual student needs, which could lead to better educational outcomes [20].

4.2. Limitations and Implications

This research presents certain limitations that need to be resolved in future studies. We still have a space for human involvement in some stages of the framework. Despite employing learning data and developing a primarily automated approach, critical decisions such as determining the appropriate number of clusters and defining specific learning patterns still rely on human intervention. This dependence on human input could potentially introduce biases or inconsistencies in the analysis. Future studies should focus on developing more autonomous systems that can minimize the need for human involvement, possibly by integrating advanced machine-learning algorithms and artificial intelligence.

This study also acknowledges certain limitations inherent in its research design, primarily attributed to its dependence on a narrowly defined dataset. The dataset underpinning our analysis lacks the breadth and diversity necessary to ensure the broad applicability of our findings across various populations and educational settings. Additionally, the analytical methods we employed were meticulously crafted to suit our dataset's specificities, suggesting that applying these methods to different sets of data might require substantial adaptation.

The interpretative aspect of our research, particularly identifying data patterns as manifestations of self-regulatory learning behaviors, warrants a cautious approach. Such interpretations are primarily speculative, lacking empirical substantiation, and not corroborated by additional data sources such as learner self-reports or qualitative insights. This speculative nature extends to the observation that a linear approach to learning tasks might represent learners' strategic, self-regulated choices informed by their knowledge and objectives.

An additional limitation arises from the study's exclusive focus on online learning activities, particularly those within Massive Open Online Courses (MOOCs). This focus leaves the realm of offline learning behaviors and their potential impacts on academic performance unexplored. Consequently, the transferability of our conclusions to offline educational environments remains uncertain.

Despite these limitations, this research has implications in practice and research, while also offering valuable insights and advancements. From a practical standpoint, the insights into learning patterns and their correlation with educational outcomes can inform the design of teaching strategies and learning materials. Educators and curriculum designers can leverage the understanding of how different learning patterns, such as sequential or self-regulated learning, impact student performance. This can lead to more personalized and effective teaching strategies catering to student's diverse needs and learning styles. Personalized teaching strategies can actually work effectively for students who do not have self-directed learning skills [49]. In addition, this research can play a critical role in developing software to create more effective and personalized learning platforms.

5. Conclusions

This study endeavors to delineate learning behaviors within learning analytics, aiming to augment decision support in educational contexts. To this end, we introduce a structured framework comprising three distinct phases: (1) Pattern Identification, which discerns unique learning behavior patterns; (2) Pattern Exploration, which delves into the nuances of these identified patterns; and (3) Pattern Evaluation, which examines the influence of these patterns on academic outcomes. Leveraging a case study within Massive Open Online Courses (MOOCs) tailored for English as a Foreign Language (EFL) learners, we identified various learning patterns, elucidated their characteristics, and analyzed their impact on academic performance. Our findings lay the groundwork for subsequent investigations, opening new paths for inquiry in data-driven education, especially within online learning environments.

Future research will seek to extend the applicability of our framework through additional case studies across diverse disciplines, thereby enhancing its generalizability. Furthermore, we plan to integrate qualitative methodologies, such as learner self-reports, to gain deeper insights into authentic self-regulated learning behaviors. Ultimately, we aspire to adapt our framework in traditional, offline educational settings. This could involve devising strategies to bridge the gap between online and offline learning modalities, thereby broadening the scope of our framework's utility in fostering effective learning strategies across various educational landscapes.

Author Contributions: Conceptualization, M.C. and K.P.; methodology, M.C. and K.P.; formal analysis, J.K. (Jiyeon Kim) and J.K. (Juhyeon Kim); investigation, J.K. (Jiyeon Kim) and J.K. (Juhyeon Kim); data curation, J.K. (Jiyeon Kim) and J.K. (Juhyeon Kim); writing—original draft preparation, M.C. and K.P.; writing—review and editing, M.C. and K.P.; supervision, K.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant funded by Ministry of Science and ICT (No. NRF-2021R1G1A1094019), the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (RS-2022-00156215) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation), and the Research Grant of Kwangwoon University in 2021.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Dumas, M.; La Rosa, M.; Mendling, J.; Reijers, H.A. *Fundamentals of Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 2.
- Rosemann, M.; vom Brocke, J. The six core elements of business process management. In *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 105–122.
- Goar, V.K.; Yadav, N.S. Business decision making by big data analytics. *Int. J. Recent Innov. Trends Comput. Commun.* **2022**, *10*, 22–35. [\[CrossRef\]](#)
- Simon, H.A. Rational decision making in business organizations. *Am. Econ. Rev.* **1979**, *69*, 493–513.
- Liu, C.; Wang, H.; Yuan, Z. A Method for Predicting the Academic Performances of College Students Based on Education System Data. *Mathematics* **2022**, *10*, 3737. [\[CrossRef\]](#)
- Ferguson, R. Learning analytics: Drivers, developments and challenges. *Int. J. Technol. Enhanc. Learn.* **2012**, *4*, 304–317. [\[CrossRef\]](#)
- Nieto-Reyes, A.; Duque, R.; Francisci, G. A method to automate the prediction of student academic performance from early stages of the course. *Mathematics* **2021**, *9*, 2677. [\[CrossRef\]](#)
- Appana, S. A review of benefits and limitations of online learning in the context of the student, the instructor and the tenured faculty. *Int. J. E-Learn.* **2008**, *7*, 5–22.
- Sahni, J. Is Learning Analytics the Future of Online Education?: Assessing Student Engagement and Academic Performance in the Online Learning Environment. *Int. J. Emerg. Technol. Learn. (Online)* **2023**, *18*, 33. [\[CrossRef\]](#)
- Montuori, L.; Alcazar-Ortega, M.; Vargas-Salgado, C.; Alfonso-Solar, D. Learning Analytics as Data driven decision making in High Education: A case study. In Proceedings of the International Conference on Innovation, Documentation and Education 2022, Valencia, Spain, 2–7 November 2022.
- Rejikumar, G.; Aswathy Asokan, A.; Sreedharan, V.R. Impact of data-driven decision-making in Lean Six Sigma: An empirical analysis. *Total Qual. Manag. Bus. Excell.* **2020**, *31*, 279–296. [\[CrossRef\]](#)
- Awan, U.; Shamim, S.; Khan, Z.; Zia, N.U.; Shariq, S.M.; Khan, M.N. Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. *Technol. Forecast. Soc. Chang.* **2021**, *168*, 120766. [\[CrossRef\]](#)
- Lu, O.H.; Huang, J.C.; Huang, A.Y.; Yang, S.J. Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. In *Learning Analytics*; Routledge: Oxfordshire, UK, 2018; pp. 78–92.
- Cobos, R.; Ruiz-Garcia, J.C. Improving learner engagement in MOOCs using a learning intervention system: A research study in engineering education. *Comput. Appl. Eng. Educ.* **2021**, *29*, 733–749. [\[CrossRef\]](#)
- Villegas-Ch, W.; Román-Cañizares, M.; Palacios-Pacheco, X. Improvement of an online education model with the integration of machine learning and data analysis in an LMS. *Appl. Sci.* **2020**, *10*, 5371. [\[CrossRef\]](#)
- Fahd, K.; Miah, S.J.; Ahmed, K. Predicting student performance in a blended learning environment using learning management system interaction data. *Appl. Comput. Inform.* **2021**, ahead-of-print.
- Panadero, E.; Jonsson, A.; Botella, J. Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educ. Res. Rev.* **2017**, *22*, 74–98. [\[CrossRef\]](#)
- Zhu, M.; Bonk, C.J.; Doo, M.Y. Self-directed learning in MOOCs: Exploring the relationships among motivation, self-monitoring, and self-management. *Educ. Technol. Res. Dev.* **2020**, *68*, 2073–2093. [\[CrossRef\]](#)
- Juhaňák, L.; Zounek, J.; Rohlíková, L. Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Comput. Hum. Behav.* **2019**, *92*, 496–506. [\[CrossRef\]](#)
- Viberg, O.; Khalil, M.; Baars, M. Self-regulated learning and learning analytics in online learning environments: A review of empirical research. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt, Germany, 23–27 March 2020; pp. 524–533.
- Takii, K.; Flanagan, B.; Ogata, H. Efl vocabulary learning using a learning analytics-based e-book and recommender platform. In Proceedings of the 2021 International Conference on Advanced Learning Technologies (ICALT), Tartu, Estonia, 12–15 July 2021; pp. 254–256.
- Lin, C.J.; Mubarak, H. Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educ. Technol. Soc.* **2021**, *24*, 16–35.
- Lin, C.J.; Hwang, G.J. A learning analytics approach to investigating factors affecting EFL students' oral performance in a flipped classroom. *J. Educ. Technol. Soc.* **2018**, *21*, 205–219.
- Murtagh, F.; Contreras, P. Algorithms for hierarchical clustering: An overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 86–97. [\[CrossRef\]](#)
- Augusto, A.; Conforti, R.; Dumas, M.; Maggi, F.M.; Marrella, A.; Mecella, M.; Soo, A. Automated discovery of process models from event logs: Review and benchmark. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 686–705. [\[CrossRef\]](#)
- Park, G.; Cho, M.; Lee, J. Leveraging machine learning for automatic topic discovery and forecasting of process mining research: A literature review. *Expert Syst. Appl.* **2023**, *239*, 122435. [\[CrossRef\]](#)
- van der Aalst, W.M. Object-Centric Process Mining: Unraveling the Fabric of Real Processes. *Mathematics* **2023**, *11*, 2691. [\[CrossRef\]](#)
- Kendall, M.G. *Rank Correlation Methods*; C. Griffin: Oxford, UK, 1948.
- Gilbert, R.O. *Statistical Methods for Environmental Pollution Monitoring*; John Wiley & Sons: Hoboken, NJ, USA, 1987.

30. Yujian, L.; Bo, L. A normalized Levenshtein distance metric. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1091–1095. [\[CrossRef\]](#)
31. Leemans, S.J.; Fahland, D.; Van Der Aalst, W.M. Discovering block-structured process models from event logs-a constructive approach. In Proceedings of the Application and Theory of Petri Nets and Concurrency: 34th International Conference, PETRI NETS 2013, Milan, Italy, 24–28 June 2013; Proceedings 34; Springer: Berlin/Heidelberg, Germany, 2013; pp. 311–329.
32. Momeni, A.; Pincus, M.; Libien, J.; Momeni, A.; Pincus, M.; Libien, J. Cross tabulation and categorical data analysis. In *Introduction to Statistical Methods in Pathology*; Springer: Cham, Switzerland, 2018; pp. 93–120.
33. Balakrishnan, N.; Voinov, V.; Nikulin, M.S. *Chi-Squared Goodness of Fit Tests with Applications*; Academic Press: Cambridge, MA, USA, 2013.
34. Chao, A.; Gotelli, N.J.; Hsieh, T.; Sander, E.L.; Ma, K.; Colwell, R.K.; Ellison, A.M. Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **2014**, *84*, 45–67. [\[CrossRef\]](#)
35. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [\[CrossRef\]](#)
36. Caso, C.; Angeles gil, M. The Gini-Simpson index of diversity: Estimation in the stratified sampling. *Commun. Stat.-Theory Methods* **1988**, *17*, 2981–2995. [\[CrossRef\]](#)
37. Morris, E.K.; Caruso, T.; Buscot, F.; Fischer, M.; Hancock, C.; Maier, T.S.; Meiners, T.; Müller, C.; Obermaier, E.; Prati, D.; et al. Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* **2014**, *4*, 3514–3524. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Ruxton, G.D. The unequal variance *t*-test is an underused alternative to Student's *t*-test and the Mann–Whitney U test. *Behav. Ecol.* **2006**, *17*, 688–690. [\[CrossRef\]](#)
39. Welch, B.L. On the comparison of several mean values: An alternative approach. *Biometrika* **1951**, *38*, 330–336. [\[CrossRef\]](#)
40. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [\[CrossRef\]](#)
41. Fournier-Viger, P.; Lin, J.C.W.; Kiran, R.U.; Koh, Y.S.; Thomas, R. A survey of sequential pattern mining. *Data Sci. Pattern Recognit.* **2017**, *1*, 54–77.
42. Günther, C.W.; Rozinat, A. Disco: Discover Your Processes. *BPM (Demos)* **2012**, *940*, 40–44.
43. Fatahi, S.; Shabanali-Fami, F.; Moradi, H. An empirical study of using sequential behavior pattern mining approach to predict learning styles. *Educ. Inf. Technol.* **2018**, *23*, 1427–1445. [\[CrossRef\]](#)
44. Ye, Z.; Jiang, L.; Li, Y.; Wang, Z.; Zhang, G.; Chen, H. Analysis of Differences in Self-Regulated Learning Behavior Patterns of Online Learners. *Electronics* **2022**, *11*, 4013. [\[CrossRef\]](#)
45. Dobashi, K.; Ho, C.P.; Fulford, C.P.; Lin, M.F.G.; Higa, C. Learning pattern classification using moodle logs and the visualization of browsing processes by time-series cross-section. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100105. [\[CrossRef\]](#)
46. Pelánek, R.; Effenberger, T.; Čechák, J. Complexity and difficulty of items in learning systems. *Int. J. Artif. Intell. Educ.* **2022**, *32*, 196–232. [\[CrossRef\]](#)
47. Lee, Y.H.; Hsiao, C.; Ho, C.H. The effects of various multimedia instructional materials on students' learning responses and outcomes: A comparative experimental study. *Comput. Hum. Behav.* **2014**, *40*, 119–132. [\[CrossRef\]](#)
48. Zimmerman, B.J. Theories of self-regulated learning and academic achievement: An overview and analysis. In *Self-Regulated Learning and Academic Achievement*; Routledge: Abingdon, UK, 2013; pp. 1–36.
49. Palacios Hidalgo, F.J.; Huertas Abril, C.A.; Gómez Parra, M.E. MOOCs: Origins, concept and didactic applications: A systematic review of the literature (2012–2019). *Technol. Knowl. Learn.* **2020**, *25*, 853–879. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.