Online Journal of Communication and Media Technologies, 2024, **14**(4), e202446
e-ISSN: 1986-3497

OPEN ACCESS

Research Article

Exploration of the online learners' actions: A sequence mining approach

Rusada Natthaphatwirata [1]
0009-0007-4728-2462

Wannisa Matcha [2*]
0000-0003-1120-3324

[1] Faculty of Education, Prince of Songkla University Pattani Campus, Pattani, THAILAND
[2] Faculty of Communication Sciences, Prince of Songkla University Pattani Campus, Pattani, THAILAND
[*] Corresponding author: wannisa.ma@psu.ac.th

Citation: Natthaphatwirata, R., & Matcha, W. (2024). Exploration of the online learners' actions: A sequence mining approach. *Online Journal of Communication and Media Technologies, 14*(4), e202446. https://doi.org/10.30935/ojcmt/14957

ARTICLE INFO

Received: 12 May 2024

Accepted: 21 Jul 2024

ABSTRACT

This paper presents the exploration of the learners' learning engagement in a self-paced massive open online course (MOOC). Research often claims that engagement contributes to learning success. However, there is still limited understanding of engagement and its characteristics. This research aims to fulfil this gap by exploring how different patterns detected based on the density levels of engagement contribute to learning performance. A total number of 159,804 records of trace data from 971 learners who enrolled in a self-paced MOOC were used in this study. The sequence mining technique was used to formulate the sequence of learning engagement. Hierarchical clustering was then used to automate the pattern recognition of the formulated sequences. As a result, four groups of learners were detected based on a similar pattern of engagement levels. Sequence mining was then used to examine the learning engagement pattern. The Kruskal-Wallis test was used to examine the statistically significant differences in terms of final scores among the detected groups. The results revealed two successful groups of learners with different patterns of engagement and two unsuccessful groups. Successful learners are intensively engaged in learning activities in the short and long run, whereas unsuccessful groups tend to be less engaged. This paper extends the previous exploration of the engagement. That is, the level identified based on the density of interactive engagement as recorded in the system can be used to determine the learning patterns, consequently, reflective of individual's learning profiles. It has a significant association with academic performance.

Keywords: learning analytics, self-regulated learning, educational data mining, sequence mining, learners' profile, MOOCs

INTRODUCTION

Massive open online courses (MOOCs) are considered the largest educational platform that are administered by many prestigious universities to offer a variety of online courses (Guo & Reinecke, 2014). The most frequently used approach to design and conduct the courses in MOOCs is self-paced learning. Self-paced learning is the general term used to explain the course that gives full control over the schedule of learning to the learners. That is, learners interact with the learning courses at their own pace and time (Vilkova, 2022). This type of MOOC provides much flexibility and benefits for learners. This is because the learning materials offered in MOOCs can be freely accessed (Wong et al., 2019), there is less time constraint in learning (Vilkova, 2022), learners can review the learning materials and revisit the lecture as many times as they want, and many more. Due to the flexibility of attending online courses, a variety of individual learners registered for MOOCs. However, not all registered learners complete the courses (Eriksson et al., 2017; Kizilcec et al., 2016; Reich & Ruipérez-Valiente, 2019). The ability to complete online courses is subject to several factors such as the learning goal, self-regulation skills (Maldonado-Mahauad et al., 2018a, 2018b), time management skills (Ahmad Uzir et al., 2019; Kizilcec et al., 2016), prior knowledge, and learning engagement (Ogunyemi et al., 2022).

Learning engagement has been reported to have a significant impact on student achievement, motivation, and retention (Froiland & Worrell, 2016; Martin & Bolliger, 2018; Olivier et al., 2019). Engagement refers to the extent to which learners pay attention, participate in learning, and interact with the learning contents (Schnitzler et al., 2021). The level of engagement is often reported as a strong contribution to the success in learning (Gillett-Swan, 2017; Järvelä et al., 2021), yet, to what extent does learning engagement contribute to the learning outcome, and how do different levels of engagement contribute to the differences in learning achievement remain the challenge in educational research. To address these gaps, the primary reason to undertake this research is to examine whether varying density levels of engagement contribute to the success of learning in the MOOC.

Previous research examined the learning engagement by using self-report instruments such as using questionnaires. Recent work of Akhuseyinoglu and Brusilovsky (2022) reported that behavioural data recorded by the system can be used to understand the learning process. Comparing the use of self-reported instruments and behavioural data, they found that a model derived from such behavioural data is much better at predicting the learning process such as performance and engagement. These findings are well-aligned with the previous work of Zhou and Winne (2012) who discovered that trace data collected by the system is predictive of goal-oriented learning behaviours. Research also claimed that trace data is a clear reflection of actual learning behaviours (Srivastava et al., 2022; Zhou & Winne, 2012). It minimises the individual's discrepancy (Jovanovic et al., 2017) and bias (Järvelä et al., 2021). Given the positive findings observed by these researchers, this study aims to investigate if the individual learners' profiles could be extracted from the automated collected trace data given the differences in their learning behaviours. Also, this study will examine if the detected learners' profiles are associated with academic performance.

LITERATURE REVIEW

Learning Engagement and Learning Profiles

Engagement is an important factor that contributes to the success of learning. It is generally perceived as the affordance of learners to complete the learning tasks (Salas- Pilco et al., 2022). Engagement is reflective of learners' interest and active participation in learning (Azevedo, 2015). Some research considered engagement as an observable indicator of intrinsic motivation (Froiland & Worrell, 2016). That is, driven by natural willingness, learners decided to participate and sustain in the learning activities designed by the course instructors. Hence, intrinsic motivation is evidenced through the behaviours that align with the tasks assigned within learning activities such as reading, collaborating with fellow learners, teamwork, conducting independent research, and more. These activities constitute the elements of the learning interaction between learners and the learning environment, fostering interest and sustained attention (Bond et al., 2020). As presented thus far, engagement can be observed in terms of behavioural, cognitive, and affective characteristics (Salas- Pilco et al., 2022).

Engagement has been reported to strongly support learning achievement (Gillett-Swan, 2017; Järvelä et al., 2021). For instance, van Rooij et al. (2017) conducted a study focusing on the transition from secondary school to university level by using a set of questionnaires. They found that those who showed low behavioural and cognitive engagement during secondary school were the least successful students when attending the university level. Those who obtained a high engagement score in secondary school performed better when entering the university level. Similarly, Schnitzler et al. (2021) conducted a study among 397 high school students in a real classroom setting. They use pre- and post-tests as well as video recordings of learning activities during the face-to-face classroom. They identified five groups of learning engagement profiles by using latent profile analysis. These five patterns include disengaged, compliant, silent, engaged, and busy. They found that those who exhibited moderate to high levels of engagement obtained higher academic self-concepts as compared to those who showed lower levels of engagement (Schnitzler et al., 2021). Martin and Bolliger (2018) studied the students' perception of engagement strategies. They found that engagement between learner and instructor is the most valuable engagement. However, research posited that engaging in online courses is much more challenging than the traditional face-to-face classroom (Gillett-Swan, 2017; Hew, 2016). This is due to the nature of online courses which offer to be learned at students' own flexibility. Hence, a lack of teachers' or instructors' presence is expected. In some cases, the feedback to motivate consistency in engagement is missing.

Nonetheless, the use of online learning technology enables automated records of learning that would enable researchers and academics to examine several dimensions of learning and engagement. This recorded data is known as trace data or log data. Trace data is the digital footprints that learners leave behind while interacting with online learning content. Trace data contains rich information about learners, interactions, actions, and time they spent engaged in learning. Hence, it can be used to examine the behavioural engagement. Several contemporary research studies make use of such data to examine learners' behavioural engagement. This behavioural engagement was previously studied in terms of the approaches of interaction (Jovanovic et al., 2017), the intensity of activities interaction (Lust et al., 2013), the sequence of actions, the event of actions (Ahmad Uziret al., 2019; Fan et al., 2022), and the process of interaction (Saint et al., 2022).

For example, van den Beemt et al. (2018) analysed the trace data by using a process mining technique to observe the insights on the learning behaviours and their relation to learning progress. Their analysis was built based on the constructivism theory. They applied the process mining and clustering technique to automate the detection of students' learning patterns. They found that learners behave differently when interacting with online courses. That is, learners with steady learning behaviours are more likely to achieve the course learning outcome. Saqr et al. (2023) conducted a longitudinal study to explore the students' engagement over four years. They used the trace data recorded by the learning management system (LMS) to explore the engagement over the course of undergraduate students. The results showed that longitudinal high levels of engagement were associated with higher academic performance. However, engagement should not be considered as a single point in time since engagement has variations. Yet, disengagement at any point in time was associated with lower achievement. Hence, some interventions or strategies need to be applied in order to promote the improvement of learning engagement.

As presented thus far, trace data contained insightful information that could be used to understand learning and its process. Given the insight derived from trace data, several research questions that were previously hard to explore can be answered. However, according to Paulsen and McCormick (2020), research exploring online learning engagement is still scarce. There is a vague understanding of engagement characteristics and how to improve engagement in online learning (Salas- Pilco et al., 2022). Previous research examined the online learning engagement by examining it in term of different types of interactive activities (Jovanovic et al., 2017; Saqr et al. 2023; van den Beemt et al., 2018), therefore, there were a novel of research suggested that types of interactive activities contribute to the learning achievement. However, there is a deficiency reporting the impact of engagement density, irrespective of activity type, on learning outcomes. Besides that, the changes of learners as individual entities often open new challenges for educators and researchers. For instance, the use of the self-paced learning design of MOOCs enables learners to flexibly interact with the learning content. Hence, they do not need to follow the course structure. Learners can skip the already-knew part of the learning content. This creates new challenges for educators. One of the questions that require an answer is how to design a course that is suitable for different groups of learners. To answer this question, understanding the learners' differences and profiling them into clusters would be the beginning (Akhuseyinoglu & Brusilovsky, 2022).

Massive Open Online Courses (MOOCs)

MOOCs are free, publicly accessible online courses that are structured in a way that learners can direct their own learning. The courses usually contain a series of video lectures, multiple types of quizzes, and a set of discussion forums. Due to its online nature, and less visibility of instructors and learning guidelines, the learners who study in an online course are often expected to obtain a particular set of skills necessary to regulate and direct their own learning. For example, Kizilcec et al. (2016) highlighted that to be successful in MOOC, learners need to acquire self-regulated learning (SRL) skills.

Unlike the traditional face-to-face classroom, most MOOCs offer courses fully online. Learners can, therefore, adjust their pace of learning to their preferences. They do not need to follow the structure of the course, even though following the course structure is advised. Given these differences, it is, therefore, necessary to investigate how individuals behave when learning online.

Maldonado-Mahauad et al. (2018b) used a set of questionnaires to collect the self-regulated learning skills of Spanish MOOC learners. They detected three clusters of learners' profiles based on the questionnaire. Based on the initial finding, they examined how each group of learners behaved when interacting with the online materials by using process mining to analyse the trace data. As a result, they found that:

- Sampling learners were those who trail the course, underperformed, and were not goal-oriented learners.
- Comprehensive learners were those who learn by following the course structure.
- Targeting learners were those who strategically select specific course content to engage with to pass the assessment.

Even though this research presented an interesting profile of learners, they utilised the questionnaire to detect the learners' profile. The questionnaire, which is one type of self-reporting instrument, is considered a reflection of an individual's perception, not the actual behaviours. Hence, recent research uses the actual learning action and examines if such profiles could be detected from the recorded trace data.

For instance, Jovanovic et al. (2017) examined the learning strategies used by students who participated in a flipped classroom. They applied sequence mining and clustering techniques to automate pattern recognition based on the learners' online actions. As a result, they found five behavioural engagement patterns that were aligned with Biggs' (1987) approaches to learning including:

- **Deep learners:** Jovanovic et al. (2017) detected two groups of students (i.e., intensive and strategic groups) who were highly active students and utilised several learning activities that were reflective of the deep learning approach.
- **Surface learners:** They detected two groups of students (namely, selective students, and highly selective students). They were characterised by emphasising on the compulsory tasks, i.e., the learning tasks that were graded. They put in a minimum effort to pass the course.
- **Strategic learners:** They detected a group of students (i.e., highly strategic group) represented by the behaviour of strategically selected learning strategies to use in a particular learning time.

Besides learning strategies, the level of engagement is often reported as a strong contribution to the success of learning (Fincham et al., 2018; Jovanovic et al., 2017). Kizilcec et al. (2013) used trace data to calculate the weekly engagement trajectories. They identified four groups of learner profiles, including:

- Completing referred to those who were able to complete a majority of the assessment activities.
- Auditing was those who simply watched the video and discarded the assessment activities.
- Disengaging was those who initially focused on assessments but constantly disengaged after the 1-3 learning topics.
- Sampling was those who simply watched a few video lectures and withdrew from the course.

As presented thus far, trace data recorded learning interaction can be used to identify the learning profiles. They discovered interesting insights from trace data by exploring the engagement relevant to the type of learning activities. However, previous research studies examined the engagement together with the type of learning activities they interacted with. Not much research investigated if the level of density and the distribution of learning engagement such as the work of Schnitzler et al. (2021) which was conducted in a face-to-face setting, could be used to detect the profile of learners who participated in MOOCs. Exploring this different angle could expand the understanding of the extent the engagement levels impact the learning process and learning performance. Therefore, two research questions were formulated to guide this study, including:

RQ1: Can we detect the distinct groups of learners based on the different density levels of learning engagement of those who enrolled in a self-paced MOOC?

RQ2: Were there any statistically significant differences between different groups of self-paced learners and the course performance?

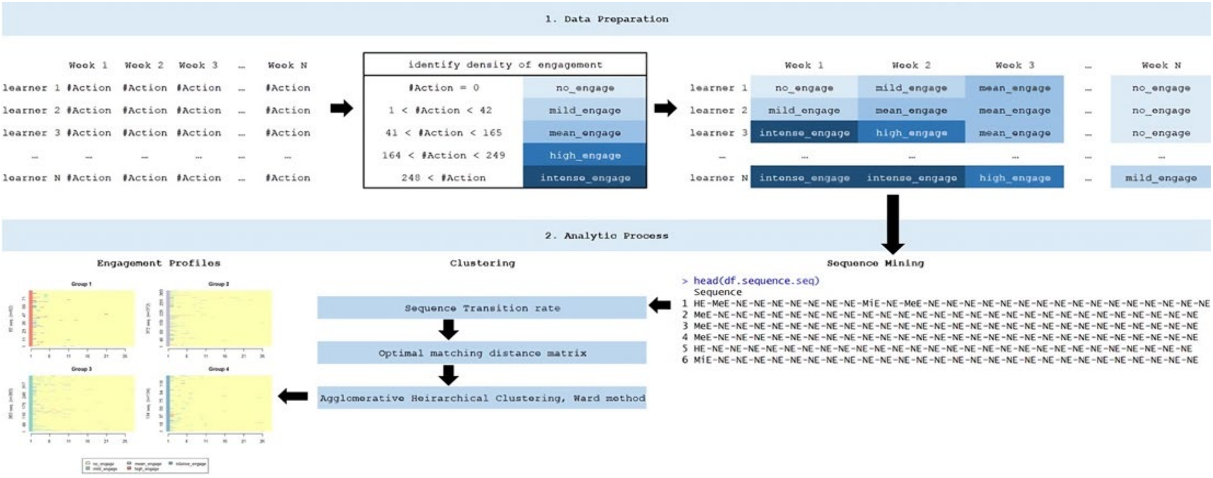


Figure 1. Data analysis process (Source: Authors)

METHOD

Learning Context

The data used for this study were collected from a MOOC course offered by a university in the Thai language. The course started to offer self-paced learning started from 2018. However, the data used in this study were collected for 2019. This short course was offered for self-paced learning which allowed learners to manage their own time to interact with the learning contents. The course focused on the fundamental of photographs, which are divided into five learning topics. In each learning topic, learners were provided with short video lectures hosted on the university's YouTube channel, reading materials, quizzes, pre-test, and post-test questionnaires, exam, and a discussion forum. They also were provided with a dashboard to monitor their progress. Learners were expected to spend about 2-3 hours per week. They need to pass 80% of assessment activities to pass the course. The LMS was used to record the learners' footprints when interacted with online content. In this study, a series of events including ID, timestamp, types of events, and types of contexts were used. The ID was anonymised to ensure the protection of learners' identity. Trace data collected from 971 learners were used in this study.

Data Analysis Method

Data analysis involves a preparation and analytic process as presented in **Figure 1**. During preparation, the data were explored by using descriptive statistics. To extract the period of time that learners begin to interact with the course, the lubridate R library was used. The timestamp collected by LMS allows us to extract the started date and also the lubridate R package allows us to translate the dates into the International Organization for Standardization (ISO) standard week. Knowing the starting week for each learner, the accumulated number of actions that learners spend interacting with the learning content in each week can be investigated. To our knowledge, there is no benchmark on the density of learning engagement. Hence, this research takes the initiative step to explore the extent to which the level of learning engagement contributes to learning performance by considering the number of actions. The number of actions per week was coded to identify the state of learning engagement based on the overall frequency of recorded actions. That is, by calculating the descriptive statistic using mean, median, and quartiles, the density levels of engagement are as follows:

- **no_engage** refers to the total zero learning actions recorded by the system.
- **mild_engage** refers to a minimal number of actions observed during that particular week, i.e., between 1 to 41 recorded learning actions (i.e., the first quartile based on frequency of recorded actions).
- **mean_engage** refers to the mean number of actions observed during that particular week, i.e., between 42 to 164 recorded learning actions.
- **high_engage** refers to a much higher number of actions observed during that particular week, i.e., between 165 to 249 recorded learning actions.

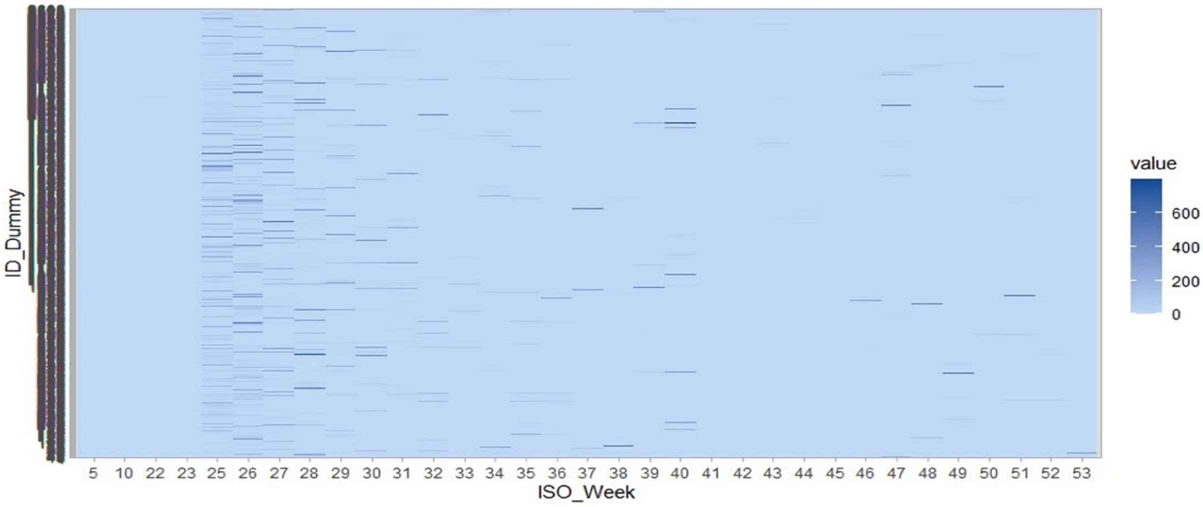


Figure 2. Overall time spent in each week in 2019 based on the international standard week (Source: Authors)

- **intense_engage** refers to an intensive afford as reflected by a high number of actions observed during that particular week, i.e., above 250 recorded learning actions.

Sequence mining was then used to visualise the sequence of actions based on the density of actions. TraMineR R library is a well-developed and frequently used R package used to create the sequence (Gabadinho et al., 2008, 2011). Sequence mining arranges the state (engagement density) and also calculates the transition between these states (Gabadinho et al., 2011). TraMineR R library also provides a function to identify the sub-sequences that are significantly distinct in each cluster by using the chi-square test. After that, the optimal matching distance matrix was calculated based on the transition rate. This matrix was then used as an input to the clustering algorithm. The Agglomerative Hierarchical clustering based on Ward method was applied to automate the grouping of learners. That is, based on the density of actions each learner performed in each learning week, the hierarchical clustering enables us to detect similar patterns of these behaviours.

The Kruskal-Wallis test was then used to examine if there were any significant differences in terms of learning performance identified by using scores in the final exam among the detected clusters. To ensure the suitability of Kruskal-Wallis test, following assumption were examined,

i) there were more than two independent groups,

ii) sample size in each independent group were more than five, and

iii) the data were non-normal distribution.

The effect size was calculated based on the Kruskal-Wallis H-statistic to examine the magnitude of the differences. Since the Kruskal-Wallis test has a limitation in detecting which groups are different from each other. It only enables to detect if at least one group is different and it does not take into account the magnitude of the difference, hence another statistical is needed. Since the data is non normal distributed, the pairwise Wilcoxon's test by using the Bonferroni p-adjustment method was computed the investigate the significant differences between each pair of clusters in terms of the learners' course performance (i.e., exam score). Bonferroni p-adjustment method is used to minimise the likelihood of obtaining false-positive results (type I errors) when multiple pairwise tests are performed on a single set of data.

RESULT

RQ1: Detection of Self-Paced Learners

The trace data collected from 971 learners who first enrolled in this MOOC in 2019 consisted of 159,804 records. As presented in **Figure 2**, the course received much attention from learners during the 24^{th} week based on the ISO calendar, which was from June 10, 2019, onward. This timeline is aligned with the start of the academic year in Thailand's educational system. Since this course was designed based on self-paced learning, hence, learners are allowed to enrol in the course at any point in time. The management of the

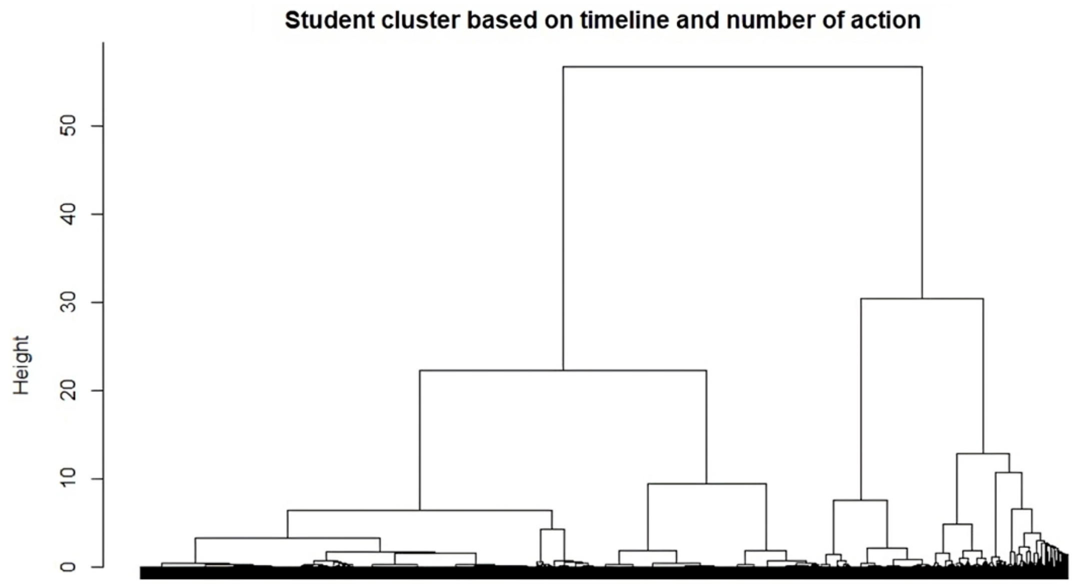
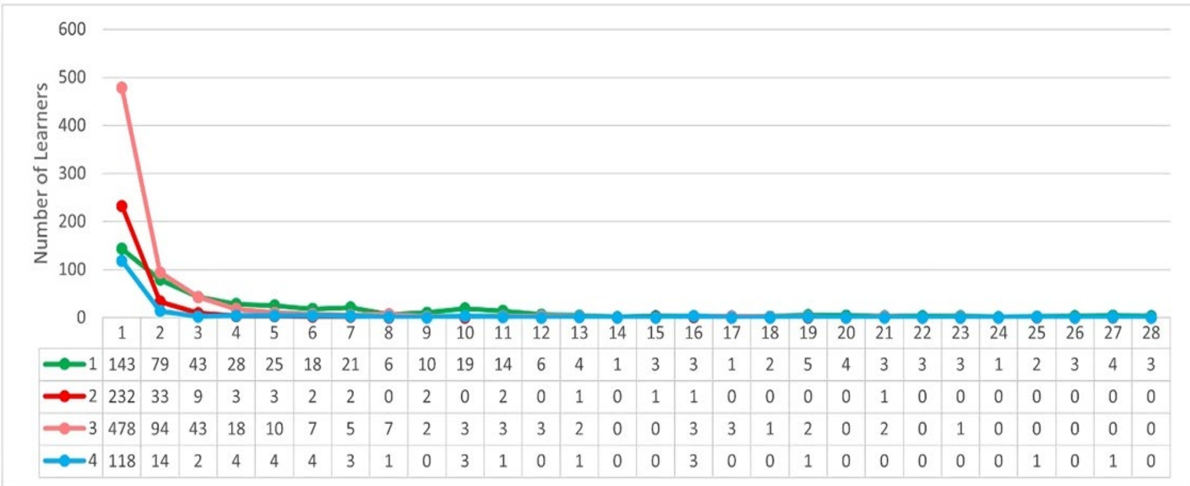


Figure 3. The dendrogram computed from the learning engagement levels of each learner (Source: Authors)



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	143	79	43	28	25	18	21	6	10	19	14	6	4	1	3	3	1	2	5	4	3	3	3	1	2	3	4	3
2	232	33	9	3	3	2	2	0	2	0	2	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
3	478	94	43	18	10	7	5	7	2	3	3	3	2	0	0	3	3	1	2	0	2	0	1	0	0	0	0	0
4	118	14	2	4	4	4	3	1	0	3	1	0	1	0	0	3	0	0	1	0	0	0	0	0	1	0	1	0

Figure 4. Number of active learners in each cluster in each week after the first enrolment to the course (Source: Authors)

learning schedule was flexible. That is, learners can spend time as much as they want to interact with the course. Based on this notion, the timeline was recalculated by considering the first-time enrolment of individual learners as week 1 for the learning schedule.

As a result of this recalculation, the shortest time spent interacting with the learning materials was one week and the longest one was 28 weeks. The number of interactions recorded in the system showed that the maximum number of interactions per week per learner was 903 learning actions, and the lowest was 10 actions. Based on the active learning time, i.e., the time learners visited the course, the system recorded about 165 learning actions per learner (median (Q1, Q3) = 93 (41, 249)).

Hierarchical clustering was then used to automate the pattern detection based on learners' time spent interacting with the learning course, i.e., the actual number of recorded actions. The dendrogram (**Figure 3**) computed from learning actions for the hierarchical clustering suggests four groups of learners. **Figure 4** presents the number of active learners who visited the course in each cluster with respect to the week.

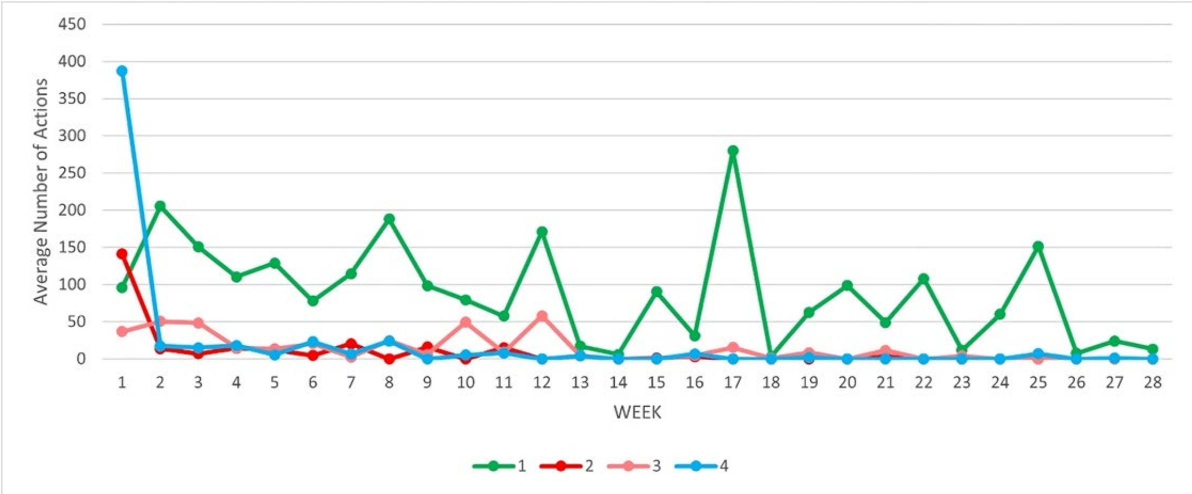


Figure 5. Average number of actions per weekly active learners in each cluster (Source: Authors)

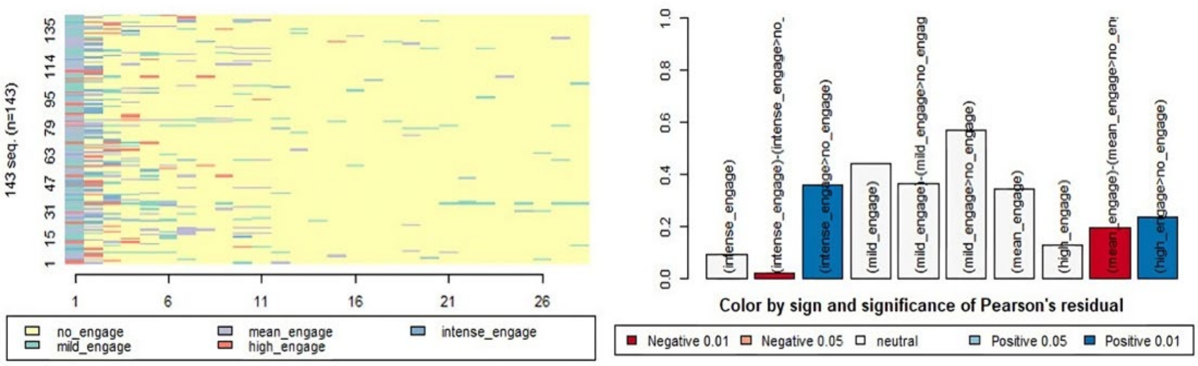


Figure 6. Transition of learning afford of Group 1 (Source: Authors)

Figure 5 shows the overall patterns of each detected cluster. To gain a better understanding of each cluster's behaviours sequence mining was used by using the corresponding amount of engagement explained earlier.

Group 1 – Distributed engaged learners

143 learners were grouped in this cluster. As presented in **Figure 6**, learners in this group put high to afford to accomplish the course. That is, the rating total numbers of actions ranged from mean, high, and intensive engagement at the beginning of the course (week 1 and week 2). Even though the number of active learners and actions dropped after the second week of enrolment, yet the continuation of learning can be seen. In terms of discriminating sub-sequence (refer to **Figure 6**, right-hand side), the transition between intensively or highly engaged in learning to no engagement at all in the following week (*intense_engage–no_engage* and *high_engage–no_engage*) was significantly distinct action as compared to other types of engagement transition.

Group 2 – Moderate engaged learners

This is considered the second largest group of learners (*N* = 232 learners). Most of the learners interacted with the course during the first week of learning. A small number of interactions in the following weeks can be seen. The discriminating sub-sequence as presented in **Figure 7** on the right-hand side of the graph showed that the transition between intensively or highly engaged in learning to no engagement at all in the following week (*mean_engage–no_engage* and *high_engage–no_engage*) were significantly positive discriminating action as compared to other types of engagement transition. Meanwhile, all sub-sequences relevant to *intense_engage* were negatively significant actions. That is, there was no *intense_engage* observed in this group.

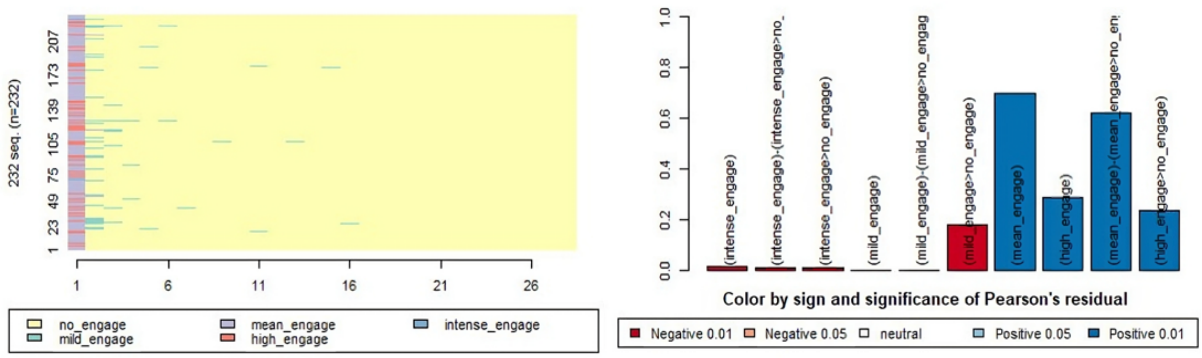


Figure 7. Transition of learning afford of Group 2 (Source: Authors)

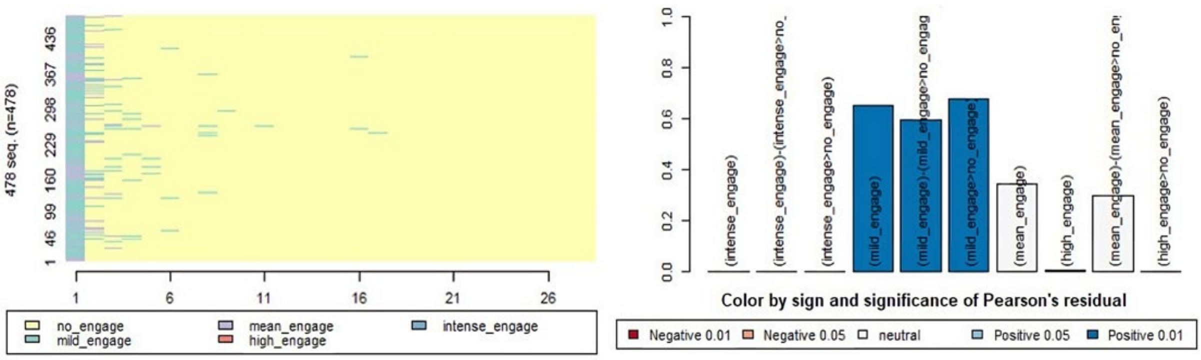


Figure 8. Transition of learning afford of Group 3 (Source: Authors)

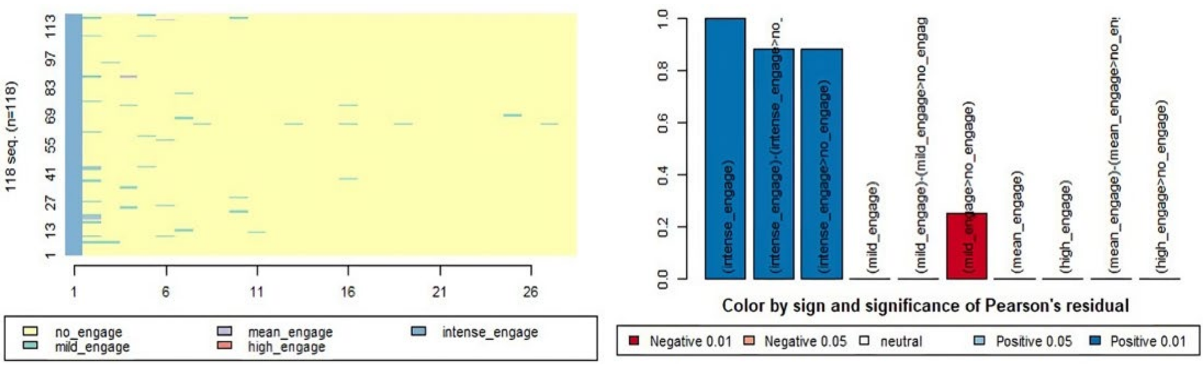


Figure 9. Transition of learning afford of Group 4 (Source: Authors)

Group 3 – Trial learners

This is the largest group of learners (*N* = 478 learners). Most of the learners in this group put a minimal effort (*mild_engage*) to participate in the learning course. Also, many of them discontinue the course after the first week. Based on their level of engagement in **Figure 8** (*mild_engage, mild_engage–no_engage*), it can be assumed that they would drop out after exploring the course for a short while.

Group 4 – Fast track learners

118 learners were categorised as *fast track learners*. As shown in **Figure 9**, learners put an intensive effort into completing the course. Most of the learners completed the course during the first week of learning. Only a few revisited the course in the following weeks. The discriminating sub-sequence as presented on the right-hand side of **Figure 9** showed that the significant positive transition was *intense_engage; intense_engage–intense_engage–no_engage; and intense_engage–no_engage*. Meanwhile, the negatively significant sub-sequence was *mild_engage–no_engage* indicating that learners in this group intensively engaged in learning activities but only for one or two weeks.

Table 1. Summary statistics for exam score for each group of learners

Groups	N	Exam (median (Q1, Q3))
Group 1–Distributed engaged learners	143	16 (0, 18)
Group 2–Moderate engaged learners	232	0 (0, 0)
Group 3–Trial learners	478	0 (0, 0)
Group 4–Fast track learners	118	17 (15, 19)

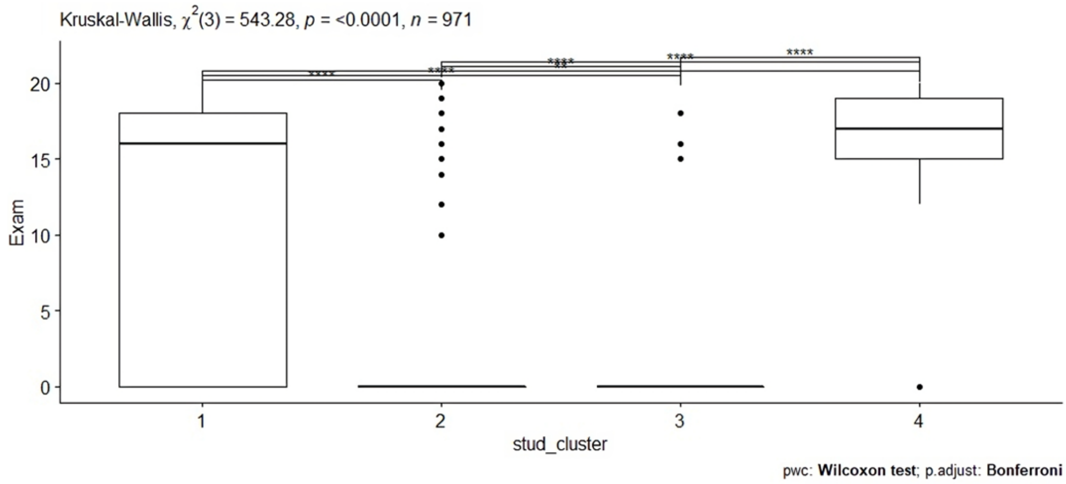


Figure 10. Kruskal-Wallis test and pairwise Wilcoxon test results (Source: Authors)

Table 2. Comparing each pair of engagement profiles in terms of exam score by using pairwise Wilcoxon test

Learning profiles		N_1	N_2	p-value	Effect size (r)	
Group 1–Distributed engaged learners	Group 2–Moderate engaged learners	143	232	0.000*	0.447	Moderate
Group 1–Distributed engaged learners	Group 3–Trial learners	143	478	0.000*	0.784	Large
Group 1–Distributed engaged learners	Group 4–Fast track learners	143	118	0.000*	0.211	Small
Group 2–Moderate engaged learners	Group 3–Trial learners	232	478	0.000*	0.380	Moderate
Group 2–Moderate engaged learners	Group 4–Fast track learners	232	118	0.000*	0.620	Large
Group 3–Trial learners	Group 4–Fast track learners	478	118	0.000*	0.930	Large

RQ2: Association with Performance

Table 1 presents data on exam performance among four distinct groups of learners, characterised by their engagement levels as presented earlier. Group 1 – distributed engaged learners, achieved a median exam score of 16, with the interquartile range (Q1, Q3) spanning from 0 to 18. Group 2 – moderate engaged learners, all recorded a median score of 0, with no variation in the interquartile range. Similarly, Group 3 – trial learners, comprises 478 members, who also had a median exam score of 0, with an identical interquartile range of 0 to 0. In contrast, Group 4 – fast track learners, attained a notably higher median score of 17, with an interquartile range from 15 to 19. These findings highlight significant differences in exam performance across the groups, correlating with their levels of engagement.

Based on the different behaviors of learners in each group, the Kruskal-Wallis test was used to examine the significant differences with respect to the course final exam among all four groups. Kruskal-Wallis reported a significant difference among the four groups of learners ($\chi^2(3) = 543.28$, p-value < 0.0001, effect size = 0.559).

The pairwise Wilcoxon's test was applied to further observe the differences between group levels by using the Bonferroni adjustment method. The statistics as presented in **Figure 10** and **Table 2**. showed significant differences between all groups of learners with respect to the final exam. The comparison between *Group 1 – distributed engaged learners* and *Group 2 – moderate engaged learners* revealed a significant difference with a p-value of 0.000 and a moderate effect size of 0.447. When comparing *Group 1 – distributed engaged learners* to *Group 3 – trial learners,* the p-value remains 0.000, and the effect size increases to 0.784, indicating a large effect. Comparing *Group 1 – distributed engaged learners* with *Group 4 – fast track learners* also results in a significant p-value of 0.000, but the effect size is smaller at 0.211.

In the comparison between *Group 2 – moderate engaged learners* and *Group 3 – trial learners*, a significant p-value of 0.000 is observed, with a moderate effect size of 0.380. When *Group 2 – moderate engaged learners* is compared to *Group 4 – fast track learners*, the p-value remains significant at 0.000, and the effect size is large at 0.620. Finally, comparing *Group 3 – trial learners* and *Group 4 – fast track learners* yields a p-value of 0.000 and a very large effect size of 0.930. These results indicate significant differences in exam scores across all pairs of engagement profiles. Almost all effect sizes of each compared group range were moderate to large effect size, except, between *Group 1 – distributed engaged learners* and *Group 4 – fast track learners.* A small difference was observed between these two groups when comparing how well they performed in the exam.

DISCUSSION

Self-paced learning design allows learners to control their own learning. There is no doubt that such a method is suitable and beneficial to the current situation, especially, when online learning has become a part of our lives. However, understanding the nature of self-paced learning has a limit. Realizing this limitation, this research work examines the profile of self-paced learning by using one of the MOOCs as a case study. Four groups of learners were identified including, *Group 1 – distributed engaged learners, Group 2 – moderate engaged learners, Group 3 – trial learners*, and *Group 4 – fast track learners*. Each group shows a different level of engagement. For instance, *Group 1 – distributed engaged learners* and *Group 4 – fast track learners* showed a high or intensive level of engagement, whereas *Group 2 – moderate engaged learners* and *Group 3 – trial learners* exhibited a lower level of engagement.

Based on the level of engagement exhibited by each group, it is no surprise that those groups who exhibited a lower level of engagement did not manage to complete the course or else obtain a lower score. Putting an effort into learning is necessary to accomplish the course (Diseth & Martinsen, 2003). However, these two groups (i.e., *Group 2 – moderate engaged learners,* and *Group 3 – trial learners*) showed a low level of engagement. *Group 3 – trial learners* showed the lowest level of engagement, they exhibited the behaviours of browsing through the content rather than looking deep into the content. This profile was previously detected by Kizilcec et al. (2013) and Maldonado-Mahauad et al. (2018b). They found a group of learners, i.e., *sampling learners*. Kizilcec et al. (2013) stated that this group of learners browsed around and watched a few videos and left or disengaged from the course. These behaviours might be explained by intrinsic motivation. After browsing through the learning content, learners might find that this course is not of their interest, hence, decided to discontinue the course. This is aligned with Maldonado-Mahauad et al. (2018b) who detected the learners' profiles based on the self-report instruments. They posited that students from the *sampling learners'* group were not driven by goal-oriented which is considered as an intrinsic motivation. Motivation is the factor that affects the learners' judgment (Ikeda, 2022), especially, in the self-paced learning course where learners register for the course of their own free will. Hence, one might have a different goal rather than completing the course. Further research into the decision to join MOOC courses might shed some light on this assumption.

Those who put a high or intensive amount of effort to engage in the course showed a significantly higher performance as demonstrated by *Group 1 – distributed engaged learners* and *Group 4 – fast track learners*. This finding corroborated with the previous works (Fincham et al., 2018; Jovanovic et al., 2017; Kizilcec et al., 2016). Previous research studies reported that the intensive amount of effort put in during the learning session has a significant association with academic performance. Research also highlighted that those who had intensive engagement in assessment and practical activities were more likely to obtain higher scores than those who focused less on the practical activities (Fincham et al., 2018; Matcha et al., 2019). Yet, the opportunity to interact with the practical activities depends on the design of the course.

In a self-paced learning course, the learning pace is expected to be unlimited. Those who are capable can complete the course as fast as they can. For instance, *Group 4 – fast track learners* were fast learners who aimed to complete the course as fast as possible. Hence, they spent the highest amount of time engaged in the learning content in the first week and completed it. Whereas *Group 1 – distributed engaged learners* spent a slightly lower level of engagement but a longer period of time to finish the course. In fact, this group of learners was representative of using the distributed learning practice. Based on Dunlosky, (2013), distributed practice refers to implementing a schedule of practice that spreads out study activities over time. It is considered one of the effective learning strategies (Dunlosky et al., 2013). Hence, they can achieve high scores on the exam. However, *Group 4 – fast track learners* showed a slightly higher score as compared to *Group 1 – distributed engaged learners. Group 1 – distributed engaged learners* consisted of a mixed score ranging from 0 to 20. Hence, future research is needed to examine the characteristics of these groups.

CONCLUSION, IMPLICATION AND LIMITATION

This research presents a study into self-paced learning MOOC learners' behaviours. Given the authority of their own learning, learners have different choices to learn. This study detected two groups of successful learners and 2 groups of unsuccessful learners. Successful learners put in a certain amount of afford to stay engaged in the course. They can either distribute their learning across a few weeks (such as *Group 1 – distributed engaged learners*) or complete it at one time (such as *Group 4 – fast track learners*). However, a certain level of effort to engage in learning must be put into practice. The unsuccess learners showed the opposite behaviour. One group showed mild engagement (*Group 3 – trial learners*). They were representative of those who were exploring and deciding whether to continue with the course or not. *Group 2 – moderate engaged learners* was different. They seemed to put some effort into interacting with the course but decided to discontinue at the end.

The findings in this research strengthen the notion that has been emphasised by many researchers that the level of engagement is an important factor in the success of learning. However, not much research explores to what extent the learning engagement contributes to the learning performance. This research uses the learning frequency as the benchmarking to determine the level of engagement. This enables to detect the distinct groups of learners who use different patterns of learning engagement. The Kruskal-Wallis test confirmed its significant association. Despite the fact observed from the statistical test, we argue that engagement solely does not explain learning success. Learning strategies used during the learning are also an important factor as discussed by the behaviours of *Group 4 – fast track learners* and *Group 1 – distributed engaged learners*. Also, unsuccess in MOOC does not necessarily identify as a failure. Many MOOC learners participated in a course with different learning goals. They might exhibit the behaviours of *Group 3 – trial learners*. Hence, designing the course by asking the learners to identify the learning goal might be a good practice. This practice might shed some light on the future design of MOOCs.

Even though the findings of this research suggest an interesting point of view. However, further exploration such as across different types or domains of MOOCs is needed in order to generalise the findings. Moreover, an exploration into each group of learners' behaviours would shed some light on the understanding, especially, in terms of their learning actions, learning strategies used, and learning process. This research suggests further research into the self-paced learning design in order to categorise and generalise the guidelines that would enhance the design of the MOOC.

Author contributions: RN: conceived of the presented idea and encapsulated the theoretical background: **WM**: designed the computational process and perform the analysis. Both authors discussed the results and contributed to the final manuscript. Both authors approved the final version of the article.

Author note: This research is a part of a project entitled: Exploration on the Knowledge Gained and MOOC Learners' Behaviors.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: This study was exempted from ethical review due to the use of historical, anonymized data and approved by the ethics committee lead by Assist. Prof. Najmee Madmam in the Research Ethics Committee for Humanities, Social Sciences and Education at Prince of Songkla University, REC number: psu.pn.2-050/65.

Declaration of interest: The authors declare no competing interest.

Data availability: Due to the university's policy and according to the research ethical agreement approved by the Prince of Songkla University, the dataset used in this study would not be made public to prevent any identifiable of the participants involved in this study.

REFERENCES

Ahmad Uzir, N., Gašević, D., Matcha, W., Jovanović, J., & Pardo, A. (2019). Analytics of time management strategies in a flipped classroom. *Journal of Computer Assisted Learning*, *36*(1), 70–88. https://doi.org/10.1111/jcal.12392

Ahmad Uzir, N., Gašević, D., Matcha, W., Jovanović, J., Pardo, A., Lim, L.-A., & Gentili, S. (2019). Discovering time management strategies in learning processes using process mining techniques. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming learning with meaningful technologies* (pp. 555–569). Springer. https://doi.org/10.1007/978-3-030-29736-7_41

Akhuseyinoglu, K., & Brusilovsky, P. (2022). Exploring behavioral patterns for data-driven modeling of learners' individual differences. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.807320

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Biggs, J. B. (1987). *Student approaches to learning and studying*. Australian Council for Educational Research.

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. *International Journal of Educational Technology in Higher Education*, *17*, Article 2. https://doi.org/10.1186/s41239-019-0176-8

Diseth, Å., & Martinsen, Ø. (2003). Approaches to learning, cognitive style, and motives as predictors of academic achievement. *Educational Psychology*, *23*(2), 195–207. https://doi.org/10.1080/01443410303225

Dunlosky, J. (2013). Strengthening the student toolbox. *American Educator*, *37*(3), 12–21.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Eriksson, T., Adawi, T., & Stöhr, C. (2017). "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education*, *29*(1), 133–146. https://doi.org/10.1007/s12528-016-9127-8

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*, *17*, 949–987. https://doi.org/10.1007/s11409-022-09291-1

Fincham, O. E., Gasevic, D. V., Jovanovic, J. M., & Pardo, A. (2018). From study tactics to learning strategies: An analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies*, *12*(1), 59–72. https://doi.org/10.1109/TLT.2018.2823317

Froiland, J. M., & Worrell, F. C. (2016). Intrinsic motivation, learning goals, engagement, and achievement in a diverse high school. *Psychology in the Schools*, *53*(3), 321–336. https://doi.org/10.1002/pits.21901

Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1–37. https://doi.org/10.18637/jss.v040.i04

Gabadinho, A., Ritschard, G., Studer, M., & Muller, N. S. (2008). *Mining sequence data in R with the TraMineR package: A user's guide*. http://mephisto.unige.ch/pub/TraMineR/doc/TraMineR-Users-Guide.pdf

Gillett-Swan, J. (2017). The challenges of online learning: Supporting and engaging the isolated learner. *Journal of Learning Design*, *10*(1), Article 20. https://doi.org/10.5204/jld.v9i3.293

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the 1st ACM Conference on Learning @ Scale Conference* (pp. 21–30). ACM. https://doi.org/10.1145/2556325.2566247

Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, *47*(2), 320–341. https://doi.org/10.1111/bjet.12235

Ikeda, K. (2022). How beliefs explain the effect of achievement goals on judgments of learning. *Metacognition and Learning*, *17*, 499–530. https://doi.org/10.1007/s11409-022-09294-y

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, *72*, Article 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Jovanovic, J., Gasevic, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*, 74–85. https://doi.org/10.1016/j.iheduc.2017.02.001

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. (2016). Recommending self-regulated learning strategies does not work (in MOOC context). In *Proceedings of the 3rd ACM Conference on Learning @Scale* (pp. 101–104). ACM. https://doi.org/10.1145/2876034.2893378

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online course. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 170–179). https://doi.org/10.1145/2460296.2460330

Lust, G., Elen, J., & Clarebout, G. (2013). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, *29*(5), 2013–2021. https://doi.org/10.1016/j.chb.2013.03.014

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

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Moreno-Marcos, P. M., Alario-Hoyos, C., Merino, P., & Delgado-Kloos, C. (2018b). Predicting learners' success in a self-paced MOOC through sequence patterns of self-regulated learning. In *Proceedings of the 13th European Conference on Technology Enhanced Learning* (pp. 355–369). https://doi.org/10.1007/978-3-319-98572-5_27

Martin, F., & Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning Journal*, *22*(1), 205–222. https://doi.org/10.24059/olj.v22i1.1092

Matcha, W., Gašević, D., Ahmad Uzir, N., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 461–470). https://doi.org/10.1145/3303772.3303787

Ogunyemi, A. A., Quaicoe, J. S., & Bauters, M. (2022). Indicators for enhancing learners' engagement in massive open online courses: A systematic review. *Computers and Education Open*, *3*, Article 100088. https://doi.org/10.1016/j.caeo.2022.100088

Olivier, E., Archambault, I., De Clercq, M., & Galand, B. (2019). Student self-efficacy, classroom engagement, and academic achievement: Comparing three theoretical frameworks. *Journal of Youth and Adolescence*, *48*(2), 326–340. https://doi.org/10.1007/s10964-018-0952-0

Paulsen, J., & McCormick, A. C. (2020). Reassessing disparities in online learner student engagement in higher education. *Educational Researcher*, *49*(1), 20–29. https://doi.org/10.3102/0013189X19898690

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, *363*(6423), 130–131. https://doi.org/10.1126/science.aav7958

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Computers and Education: Artificial Intelligence*, *3*, Article 100060. https://doi.org/10.1016/j.caeai.2022.100060

Salas- Pilco, S. Z., Yang, Y., & Zhang, Z. (2022). Student engagement in online learning in Latin American higher education during the COVID-19 pandemic: A systematic review. *British Journal of Educational Technology*, *53*(3), 593–619. https://doi.org/10.1111/bjet.13190

Saqr, M., López-Pernas, S., Helske, S., & Hrastinski, S. (2023). The longitudinal association between engagement and achievement varies by time, students' profiles, and achievement state: A full program study. *Computers and Education*, *199*, Article 104787. https://doi.org/10.1016/j.compedu.2023.104787

Schnitzler, K., Holzberger, D., & Seidel, T. (2021). All better than being disengaged: Student engagement patterns and their relations to academic self-concept and achievement. *European Journal of Psychology of Education*, *36*(3), 627–652. https://doi.org/10.1007/s10212-020-00500-6

Srivastava, N., Fan, Y., Rakovic, M., Singh, S., Jovanovic, J., Van Der Graaf, J., Lim, L., Surendrannair, S., Kilgour, J., Molenaar, I., Bannert, M., Moore, J., & Gasevic, D. (2022). Effects of internal and external conditions on strategies of self-regulated learning: A learning analytics study. In *Proceedings of the 12th International Learning Analytics and Knowledge Conference* (pp. 392–403). ACM. https://doi.org/10.1145/3506860.3506972

van den Beemt, A., Buys, J., & van der Aalst, W. (2018). Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on process mining and clustering. *International Review of Research in Open and Distance Learning*, *19*(5), 38–60. https://doi.org/10.19173/irrodl.v19i5.3748

van Rooij, E. C. M., Jansen, E. P. W. A., & van de Grift, W. J. C. M. (2017). Secondary school students' engagement profiles and their relationship with academic adjustment and achievement in university. *Learning and Individual Differences*, *54*, 9–19. https://doi.org/10.1016/j.lindif.2017.01.004

Vilkova, K. (2022). The promises and pitfalls of self-regulated learning interventions in MOOCs. *Technology, Knowledge and Learning*, *27*(3), 689–705. https://doi.org/10.1007/s10758-021-09580-9

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

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419. https://doi.org/10.1016/j.learninstruc.2012.03.004