Interactions of reading and assessment activities

Niels Seidel[1][0000−0003−1209−5038] and Dennis Menze[1][0000−0003−0002−868X]

FernUniversität in Hagen, Universitätsstraße 1, 58084 Hagen, Germany
niels.seidel@fernuni-hagen.de
dennis.menze@fernuni-hagen.de

Abstract. Reading and assessment are elementary activities for knowledge acquisition in online learning. Assessments represented as quizzes can help learners to identify gaps in their knowledge and understanding, which they can then overcome by reading the corresponding text-based course material. Reversely, quizzes can be used to evaluate reading comprehension. The predominantly self-regulated interaction of reading and quiz activities in learning systems used in higher education has been little studied. In this paper, we examine this interaction using scroll and log data from an online undergraduate course (N=142). By analyzing processes and sequential patterns in user sessions, we identified six session clusters for characteristic reading and quiz patterns potentially relevant for adaptive learning support. These clusters showed that individual user sessions included either mainly reading or quizzes, but rarely both.

Keywords: reading analytics · learning analytics · assessment · sequential pattern mining

1 Introduction

The acquisition of knowledge through reading and quizzes is a fundamental activity in online learning. Quizzes as formative assessments can help learners to identify gaps in their knowledge and understanding, which they can then overcome by reading the corresponding text-based learning material, e.g. eBooks or digital textbooks. The other way around, quizzes can be used to evaluate reading comprehension. Other than in video-based learning (e.g. MOOCs [6]) or many printed textbooks, in common online courses, the quizzes are separated from text-based knowledge acquisition. This separation is probably caused by the modular design of learning systems providing e.g. HTML pages and test environments. However, teachers invest a lot of effort to provide useful quizzes that correspond to the provided text-based learning materials. The interaction of reading and quiz activities under such conditions, in which students are free to combine activities and personalize learning paths over the semester, has hardly been investigated so far. We are using a Learning Analytics approach to analyze these individual interactions between reading and quiz activities. With this

approach, we aim to get further insights on how to adaptively support learners performing these activities at a time and over the course of a semester.

Regarding quizzes, events like attempt start, submission of solutions, and retries, among others are collected in the logs of an online learning environment, e.g. a Learning Management System. Similarly, reading activity can be indicated through scrolling events or page turns since eye tracking is not feasible in practice. To identify behaviors from these log events, sequences of events within individual user sessions have to be considered. Frequently occurring event sequences are referred to as sequential patterns. With regard to sequential patterns, we want to answer the research question (RQ1): What sequential patterns can be identified in reading and quiz activities? Using proven sequential pattern mining algorithms, we are analyzing reading and quiz activities within the individual user sessions. From these patterns, we identified clusters of frequent learning behaviors. From this analysis, we expect insights about situations that may require an adaptive learning support.

The remainder of the article is structured as follows. In section 2, we refer to related works. The applied methods are described in section 3, before we present the results and discussion in section 4 and 5. The article ends with a summary and outlook in section 6.

2   Related Works

The precise analysis of reading behavior is not very widespread in the field of learning analytics. Using eBooks, [2] distinguished sequential and responsive reading behavior in a study (N=90) in order to determine engagement per page, content, and student. Reading engagement correlated with final grade while the proposed reading styles did not. The authors did not consider assessment and changes in the reading style over time. [14] formed three groups from a cohort of 160 graduate students considering their reading motivation and reading duration with regard to four course texts. The learning behavior was coded into six sequential patterns: intensive reading, multi-tasking reading, skim-reading, passing a course unit test, not completing a test, and being inactive. Changes between these behaviors revealed differences between the three groups. After performing a test, the two groups with low reading duration bypassed reading material in favor of another test. Test performances did not correspond to the reading behavior. However, reading was roughly measured by page turns instead of paragraph views. [19] modeled scrolling activities to predict the revisitation of short text sections to define implicit bookmarks. In comparison to the aforementioned works, scroll data was precisely recorded and analyzed over time.

The relationship between reading and quiz activities in higher education has only been examined in a few studies so far. [4] analyzed respective relations using an interactive textbook. About 700 computer science students made use of mandatory quizzes and optional reading tasks. The majority of students directly performed the quizzes without reading any text. Those participants who were reading did it shortly before the due date of their homework and could not spend

much time on reading. In an experimental design (N=36/38), [18] compared reading comprehension using an eBook system with and without generated cloze items. Compared to the control group, the experimental group benefited from quiz items that effectively promoted their reading skills, reading engagement, and reading comprehension. The undergraduate participants could improve their reading comprehension through repeated tests.

In a large study (N=13,362) of 412 courses in China, [3] used transfer state diagrams to describe changes between interactions with the course material and among peers and the instructor. Longer interaction sequences have not been considered. Although the changes have been compared between three periods of one months, students have not been grouped based on behavioral similarities.

In summary, the interaction of reading and assessment has not been sufficiently explored. Especially reading analytics could be improved by precise scrolling measures. Methods like process mining and sequential pattern mining in combination with clustering have not been used to consider relations of reading and assessment.

3   Methods

3.1   Data

Participants and design: The study was conducted in the compulsory course "Operating Systems and Computer Networks" of a distance learning B.Sc. Computer Science study program (CS course) in the winter semester 2020/2021. For the enrolled students, a supplementary course was set up in a Moodle learning environment. The use of the learning environment was voluntary, but conditional on a two-step consent to use the platform and to participate in the study. As an incentive for students' participation, additional learning opportunities such as self-assessments, assignments, semester planning support, and interactive course texts have been provided. These differences in the learning offers are comparable to different didactic offers of tutors in face-to-face teaching. Students not participating in the study had no disadvantages regarding the examination since the course texts provided to all enrolled students form the basis of the examination. 180 of the 534 CS course participants agreed to take part in the study and to use the Moodle environment. By the end of the semester, the same number of active participants had been recorded, but only 142 of them used the quizzes or texts offered in the course. The participating students were between 19 and 65 years old (M=37.21, SD=9.03). 128 were male and 52 female.

Material: The Moodle course contained four units including course texts, a newsgroup forum, recordings of live sessions, 30 assignments corrected by a tutor, and questions for exam preparation. For this study, 42 self-assessment questions [5] and 23 multiple-choice questions were provided, both referred as quizzes.

Textual learning resources in the form of study letters are a core element of distance learning at the FernUniversität in Hagen. For the proposed field study,

we have converted the existing course units from Word/LaTeX format to HTML. The HTML could be used in the Moodle activity plugin called Longpage. This plugin was developed by the authors to support reading of comprehensive and interactive course materials of the size of 40 to 80 standard pages. In a reader-friendly design, basic instruments have been provided to facilitate navigation and orientation. For instance, a table of contents, individual bookmarks, cross-references, and a full-text search was included. Additional cross-references to semantically related courses were realized on the basis of text corpora [10] using a recommender system [12].

Course texts and quizzes are not embedded together on the same page, but separate, although there are some cross-links: texts contain several hyperlinks to corresponding quizzes, and the feedback to a submitted quiz can include hyperlinks to the corresponding text sections. The use of these links was not evaluated in this study.

As shown in Fig. 1, individual reading progress was graphically emphasized at the margins of text sections.
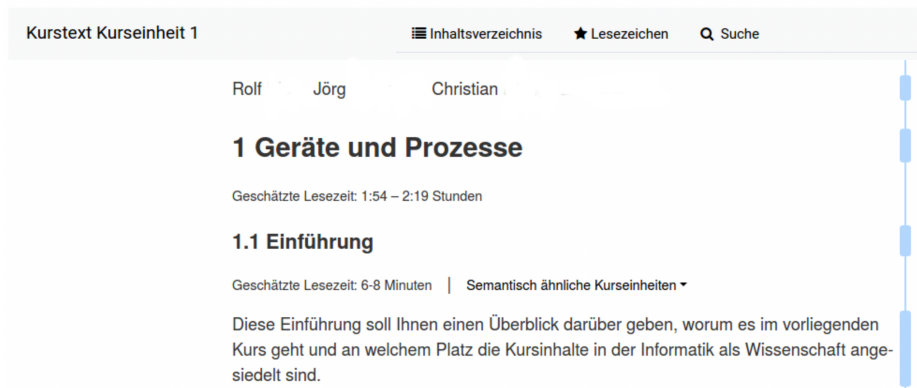
Fig. 1. Presentation of a course unit using the so called Longpage Moodle plugin.

Data collection and pre-processing: User interactions within the Moodle environment have been captured in the database, especially in the standard log store. To capture real usage data on users' reading behavior, we used the Intersection Observer API that is available in a modern web browser. The Intersection Observer fires log events as soon as a text section becomes visible within the viewport of the users' display. Text sections had a unique identifier and contained individual paragraphs, headlines, lists, images, and listings. The dataset consisted of 238,166 log entries from reading and quiz activities related to 1,359 individual user sessions. A user session is defined by a continuous sequence of consecutive log entries related to the quiz or reading activities of a user with a time difference between subsequent log entries of less than 45 minutes [7]. Fig. 2 shows an example of a reading session (being a sub-session of a user session)

indicating the beginning and end of reading sessions in comparison to shorter activity breaks. Reading sessions were categorized based on tertiles of reading durations in minutes (intervals: $(0, 1] < (1, 5] < (5, 60]$). From these activities and sessions, the events shown in Tab. 1 were derived (10,079 in total).
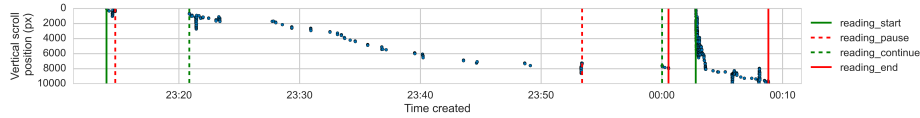
Fig. 2. Example of a reading session derived from scroll events indicating reading begin and end as well as activity breaks

Table 1. Definition of reading and quiz events within a session

Event	Description	N
reading_start	first scroll event on page after at least 10min	987
reading_short	< 1min time window of scroll events on particular text page	576
reading_medium	1-5min time window of scroll events on particular text page	266
reading_long	> 5min time window of scroll events on particular text page	355
reading_pause	< 5min break without scroll events on particular text page	210
reading_continue	first scroll event after break on particular text page	210
reading_end	> 10min break without scroll events on particular text page	987
quiz_start	first time of opening particular quiz	2,785
quiz_repeat_same	repeated quiz attempt of same quiz after success or fail	567
quiz_repeat_other	repeated quiz attempt of a quiz unrelated to the last one	313
quiz_success	submitted a solution that is more than 80 % correct (cf. [8])	1,473
quiz_fail	submitted a solution that is less than or equal to 80 % correct	1,350

3.2 Mining processes and sequences

To gain further insights into common patterns of user sessions, different methods may be used: trace or profile clustering ([13, 17]), where a trace or profile could be the vector of frequencies of events per session, or the so-called hypothesis-driven approach from [1] of finding study patterns by labeling activity sub-sequences. The latter was used in this study. Thus, from each user session, nominal features were manually generated according to the following sequential properties: whether the sequence of events per session starts with, ends with and/or mainly consists of reading or quiz activities; and in which tertile the length of the sequence of events per session belongs (intervals: $[1, 3] < (3, 7] < (7, 84]$). As there is no prior knowledge of how to best cluster the dataset, the unsupervised method k-means was used for clustering user sessions by their properties. The

number of clusters was determined by the distortion score and silhouette score ([11]). The distortion score is defined as the mean of the sum of squared distances of data points to the center of the cluster.

Each cluster, containing a subset of user sessions, was further classified as to whether it contained mainly quiz activities or reading activities, or both. Next, on each cluster, HeuristicsMiner [16] was used for visualizing the sessions as state transition charts. This method identifies state transitions between pairs of states A, B above a defined dependency threshold calculated as follows:

$$dependency(A, B) = \frac{|A \rightarrow B| - |B \rightarrow A|}{|A \rightarrow B| + |B \rightarrow A| + 1} \tag{1}$$

Since only pairwise activity sequences are considered when looking at state transitions, we also analyzed longer activity sequences using PrefixSpan algorithm [9] for mining sequential patterns. To determine the most frequent sequences with a minimal length of three activities, the Support measure has been employed. The Support is defined as the proportion between the number of sequences containing a certain sequence and the number of all sequences. It ranges from 0 to 1, where 0 means that the sequence did not occur at all and 1 that the sequence occurred in all patterns.

4   Results

After presenting very brief results from a descriptive analysis, the results of the process and sequence mining of user sessions will be presented in the second part of the section. The anonymized data and analysis scripts used are publicly available at https://anonymous.4open.science/r/A0F145 (last accessed 2022/07/01).

4.1   Descriptive Analysis

For a descriptive analysis of the quiz activities, we refer to one of our previous papers [5]. Fig. 3 shows how much of the estimated reading time students spent on the course unit (CU) texts during the whole semester. For CU1, for instance, 11% of students had reading sessions equal to or longer than 100% of the estimated CU reading time.[1] Referring to the empirical results of Trauzettel et al. [15], reading time for German language was estimated by the average number of characters read per minute minus the variance of reading speed. We assumed a reading speed lower than average because of the comparatively high text complexity of academic texts.

4.2   Mining processes and sequences

For k ranging from 2 to 20, we used a silhouette analysis [11] to find the best k representative clusters for the supplied dataset, using measurements of average

---

[1] Note that the percentage of students in Fig. 3 does not add up to 100 because less than 5% reading time is not represented.
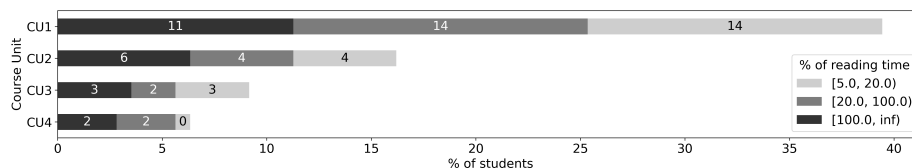
Fig. 3. Percentage of estimated reading time spent on page.

distance inside the clusters and the average distance between the clusters. For $k = 6$, the best result was obtained (silhouette score $= 0.89$, distortion score $=$ 111.63). As shown in Tab. 2, the size of the clusters is represented by the number of user sessions ranging from 94 to 394. The sessions covered by each cluster have been performed by a significant portion of participants (23.9% – 66.9%). For visualizing the transitions between the actions in a cluster, significant transitions are drawn in a transition state diagram using the HeuristicsMiner [16] with a dependency threshold of 0.9. The action sequences in every cluster were further classified as to whether they contained mainly quiz activities (SC1, SC3, and SC4) or reading activities (SC2 and SC6), or both (SC5). From the session clusters, state transition charts have been derived to support the interpretation of the data. The Fig. 4–7 show the state transition charts. The states correspond to the events listed in Tab. 1, while transitions are indicated by outgoing arrows pointing to the subsequent activity. The labels of the transition show the probability for a transition to the respective state with a range between 0 and 1. A value of 0 would indicate no transition to another state, while a value of 1 would indicate that a transition exists to a single connected state. Transition with probabilities below the threshold of 0.01 have not been displayed to improve readability.

Table 2. Session clusters (SC)

Session clusters	SC1	SC2	SC3	SC4	SC5	SC6	Total
Sessions (%)	234	207	394	261	169	94	1359
	(17.2)	(15.2)	(29.0)	(19.2)	(12.4)	(6.9)	(100)
Users (%)	82 (57.8)	71 (50.0)	95 (66.9)	71 (50.0)	34 (23.9)	48 (33.8)	142
							(100)

SC1 only contains quiz activities (Fig. 4). 22 % of those who call up a quiz continue to search other quizzes until they find one that suits them. If they fail a quiz, they continue with another one. The majority of those who accomplish a quiz successfully go for another one. Only a small fraction of 4 % (72 cases) look for mastery and repeat the same quiz in order to achieve the full score. Overall, the participants selectively go through the offered quizzes and perform some of them.
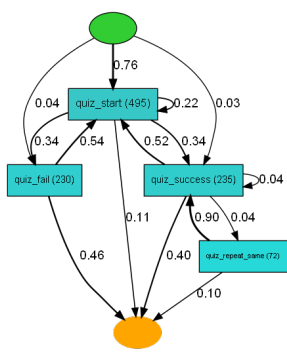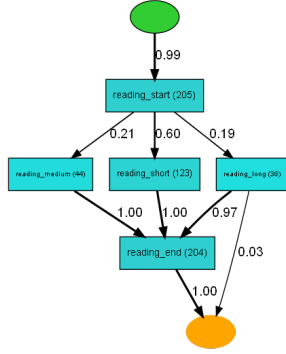
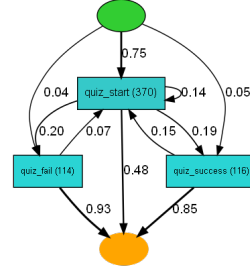**Fig. 4.** Session cluster 1   **Fig. 5.** Session cluster 2   **Fig. 6.** Session cluster 3
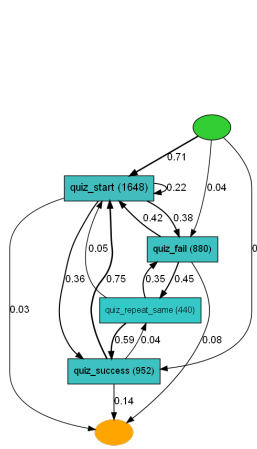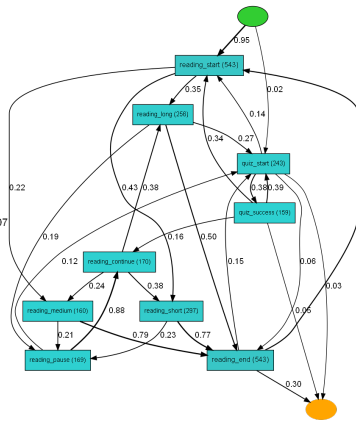

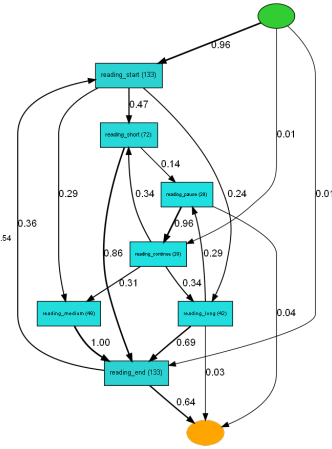**Fig. 7.** Session cluster 4   **Fig. 8.** Session cluster 5   **Fig. 9.** Session cluster 6

Reading-only activities are part of SC2 (see an example session in Fig. 2, and Fig. 5). One-fifth each starts reading for a long and medium duration of time. 60 % read only for a short time. No further findings could be derived from the three mined sequences. Since the learners do not interrupt their reading activities by pauses, another course unit, or quizzes, the activities shown in the sessions in SC2 indicate a directed and deliberate behavior shown by half of the participants.

SC3 is the largest cluster in size and the second quiz-only cluster (Fig. 6). In almost half of the sessions, attempting a quiz will terminate the session without submitting a solution. 93 % of the sessions with a failed attempt will not be continued. Only 7 % who fail go for another quiz. 15 % of the successful attempts lead to the start of another attempt, but the majority terminates the session. For this session cluster, only sequences of length two were identified. SC3 is characterized by canceling quizzes after reading the quiz description or doing

one attempt. About two-thirds of the participants showed this behavior.

SC4 as the third, shown in Fig. 7, quiz-only cluster represents almost one-fifth of the sessions. About the same percentage of learners who start an attempt are successful or not. 42 % of those who fail try another quiz, while 45 % retry the same quiz to improve. After repeating the same quiz, 59 % are successful, but 35 % fail again. Only 5 % switch to another quiz. 4 % of those who successfully complete a quiz go for mastery to become even better. Half of the participants performed sessions as described in SC4. In these sessions, they intensively try to improve their quiz performance and look forward to tackling multiple quizzes.

SC5 as represented in Fig. 8 is related to both, reading and quiz activities, but almost all sessions (95 %) start with reading for a short (43 %), medium (22 %), or long (35 %) time. Quiz attempts are made after readings end (15 %) or during reading long (27 %) or a reading pause (12 %). 38 % of the quiz attempts are successful and 39 % of those are followed by another quiz attempt. Learners in this SC do not submit solutions that let them fail. The activity quiz_fail is not present at all. Instead of risking a failed quiz, they end reading or go for another quiz. 16 % continue to read the previously used text, while 34 % go for another text after the successful completion of a quiz. SC5 is characterized by a strong interaction of reading and quiz activities.

SC6 refers to reading-only activities (Fig. 9) and describes transitions between short, medium, and long reading phases after a reading break. Breaks from reading follow 29 % of the long and 14 % of short reading phases. After a reading break, one-third of each continued reading for a short, medium, and longer time. The user sessions assembled in this session cluster are characterized by reading multiple course units within a session. No further findings could be derived from the eight mined sequences.

5   Discussion

For analysis, we applied methods like process mining, sequence mining, and clustering. In the particular context of online learning in distance education, the application of these methods and their results require a critical reflection. First of all, this analysis frames reading and quizzes as the main course activities but ignores other activities like newsgroup discussions, assignments, live sessions, and self-regulated learning support.

Session identification is an important phase in web usage mining, which can be done using time-oriented heuristics with session timeout intervals, as in our work, or navigation-oriented heuristics that consider the connection between content [7]. The latter would regard course units, chapters, etc. as logical boundaries within the course, which takes into account that sometimes a student finishes reading in one day and then resumes learning later with quiz activities. This would provide a better way to examine student behavior for somewhat independent course units. However, we preferred time-oriented heuristics because some quizzes refer to multiple course texts, and most importantly, we view a learning session as a continuous stream of consciousness within which short-term memory

is filled with temporary content to make associations between concepts for the purpose of learning. This brings the temporal aspect to the fore.

Process mining better characterizes the whole session in abstract terms, while sequence mining identifies frequent shorter sequences of the sessions. By applying both methods, we could combine the advantages of both approaches. We considered the start, end, and proportion of reading and quiz activities as nominal features for clustering user sessions. Other approaches employed the frequency of activities (trace profile) for clustering, or process mining on each user session and then clustering the dependency matrix. However, the selected features include structural characteristics that indicate the intention for initiating and terminating a session. Furthermore, we were able to consider a comparatively large set of unique activities. Although the middle part of a session has been roughly mapped as a clustering feature, we employed sequential pattern mining to describe frequent patterns.

Representative user sessions representing the most frequent reading and quiz behaviors (ignoring other types of activities which are also much less used by students) could be identified through clustering as a means of answering RQ1. Concerning quizzes, we found selective behavior to find preferred quizzes (SC1), canceling inappropriate quizzes (SC3), and intensive quiz sessions (SC4). Reading behaviors manifested in sessions with either a single course unit (SC2) or multiple course units (SC6). Also, interactions between reading and quiz activities could be observed (SC5). These behaviors can be associated with learning strategies indicating learning progress and difficulties and thus used for adaptive learning support. A prevalence of sessions that are less conducive to learning (e.g., SC1 or SC3) or a low variance in types of sessions (e.g. SC1–SC4, SC6) and thus solely reading or assessment activities could be countered by adaptive suggestions for appropriate learning strategies: students that often fall into clusters that are not beneficial for learning (e.g. skipping many quizzes, too many or few retries, etc.) could be supported by a reflection of their behavior.

The analysis underlines the importance of assessments in terms of quizzes. However, the knowledge required to successfully perform the quizzes was provided in the course texts. In the present analysis, we did not consider these semantical interrelations between quizzes and course texts.

After quizzes, the provided reading facilities proved to be important in about one-third of the sessions and for half of the participants. Although the participants tend for short reading phases, considerable medium and long reading spans could be observed. However, in the majority of user sessions, learners refuse to read course texts in Moodle. Likely, they do not refuse to read at all, since they successfully accomplished some of the quizzes. Despite of the added values provided in the Longpage plugin, they may have preferred the print or PDF versions of the course texts. Note, that the printed texts contained QR-Codes pointing directly to quizzes in the Moodle course.

A strong interplay of reading and quizzes could not be confirmed by our analysis. Although some clusters of students used quizzes and course texts within

the same user sessions, the mixture appears not as a predominant behavioral pattern concerning the user sessions.

In this study, quiz and text difficulty were not measured directly. In order to examine the extent to which certain behaviors are related to a particular course unit (perhaps some units are difficult so that students need to try the quiz again, and others are easy so that even if students fail, they think they do not need to try the quiz again) in a preliminary analysis, the correlation between the proportion of session clusters per course unit and the average correctness of answers was calculated. As a result, course units 2 and 3 were significantly more difficult than course units 1 and 4 by this measure: 2 and 3 had on average significantly about 5% fewer correct answers than 1 and 4, which was strongly negatively correlated with the relative frequency of session cluster 1 and strongly positively correlated with session cluster 5. Thus, more difficult content led to many different quizzes being taken per session, but less mixing of reading and quizzing in one session.

In a more detailed analysis, we could distinguish the paths from quiz to text and vice versa. For instance, the course texts contain several hyperlinks to corresponding quizzes, while the feedback to a submitted quiz include hyperlinks to the corresponding text sections. From the perspective of self-regulated learning, these cross-links are important and could be adaptively emphasized during corresponding reading and assessment activities.

6   Conclusion and Outlook

In this study, we aimed at analyzing interactions of reading and assessment that are potentially relevant to adaptively support learners. We could identify six session clusters of reading and quiz activities in our data (RQ1), which we further classified as comprising mainly quiz, mainly reading, or both reading and quiz activities. They showed subtle differences in their respective quiz and reading sequences: how many quizzes were tackled per session, if quizzes were repeated to deepen knowledge after a success or a fail, and how much reading per session and with or without breaks occurred. A strong relationship between reading and quiz activities per session could not be found.

In a further study, the found patterns (session clusters) should be tried to be replicated in the following semester of the same course with a different cohort. The prediction of session clusters from the first log events could be elaborated using decision trees to investigate possible interventions more concretely. The dependence between behaviors exhibited in session clusters and difficulty of the material raised in the discussion needs to be studied in more depth. Correlations of found patterns with other factors like grades, assignment results and course re-enrollment should be studied. Also, a correlation between the observed frequency of certain clusters over the course of a semester per student and their dropout rate would be important for early intervention by teachers. Furthermore, existing cross-links between quizzes and texts could be investigated. Since the winter term 2021/22, the Longpage plugin supports text highlighting and co-reading

through threaded anchored discussions and visualization of group-related reading progress. This will enable further analysis of reading behavior in terms of the expected added value for learning.

With the presented analytical approach, we hypothesized types of user session with similar reading and quiz behavior that could be used as indicators for adaptive personalized learning support. Adaptation could scaffold flexible participation and self-regulation through the close interaction of reading and quizzes.

References

[1] Boroujeni, M.S., Dillenbourg, P.: Discovery and temporal analysis of latent study patterns in MOOC interaction sequences. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. pp. 206–215. ACM, New York, NY, USA (3 2018). https://doi.org/10.1145/3170358.3170388

[2] Boticki, I., Akcapinar, G., Ogata, H.: E-book user modelling through learning analytics: the case of learner engagement and reading styles. Interactive Learning Environments **27**, 754–765 (8 2019). https://doi.org/10.1080/10494820.2019.1610459

[3] Cheng, H.N.H., Liu, Z., Sun, J., Liu, S., Yang, Z.: Unfolding online learning behavioral patterns and their temporal changes of college students in spocs. Interactive Learning Environments **25**, 176–188 (4 2017). https://doi.org/10.1080/10494820.2016.1276082

[4] Fouh, E., Breakiron, D.A., Hamouda, S., Farghally, M.F., Shaffer, C.A.: Exploring students learning behavior with an interactive etextbook in computer science courses. Computers in Human Behavior **41**, 478–485 (2014)

[5] Haake, J.M., Seidel, N., Burchart, M., Karolyi, H., Kasakowskij, R.: Accuracy of self-assessments in higher education. pp. 97–108 (2021)

[6] Kovacs, G.: Effects of in-video quizzes on MOOC lecture viewing. In: L@S 2016 - Proceedings of the 3rd 2016 ACM Conference on Learning at Scale. pp. 31–40 (2016). https://doi.org/10.1145/2876034.2876041

[7] Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R.S., Hatala, M.: Penetrating the black box of time-on-task estimation. In: ACM International Conference Proceeding Series. vol. 16-20-Marc, pp. 184–193. Association for Computing Machinery (mar 2015). https://doi.org/10.1145/2723576.2723623

[8] Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R.F., Morales, N., Munoz-Gama, J.: Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. Computers in Human Behavior **80**, 179–196 (2018). https://doi.org/10.1016/j.chb.2017.11.011

 [9] Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M.C.: PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. Proceedings - International Conference on Data Engineering pp. 215–224 (2001). https://doi.org/10.1109/icde.2001.914830

[10] Rieger, M.C., Seidel, N.: Semantic Textual Similarity von textuellen Lernmaterialien. In: Pinkwart, N., Konert, J. (eds.) Die 17. Fachtagung Bildungstechnologien, Lecture Notes in Informatics (LNI). pp. 33–44. Gesellschaft für Informatik, Bonn (2019)

[11] Rousseeuw, P.: Rousseeuw, p.j.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. comput. appl. math. 20, 53-65. Journal of Computational and Applied Mathematics **20**, 53–65 (11 1987). https://doi.org/10.1016/0377-0427(87)90125-7

[12] Seidel, N., Rieger, M.C., Walle, A.: Semantic Textual Similarity of Course Materials at a Distance-Learning University. In: Thomas W. Price, Peter Brusilovsky, Sharon I-Han Hsiao, Ken Koedinger, Shi, Y. (eds.) Proceedings of 4th Educational Data Mining in Computer Science Education (CSEDM) Workshop co-located with the 13th Educational Data Mining Conference (EDM 2020), Virtual Event, July 10, 2020. CEUR-WS.org (2020), http://ceur-ws.org/Vol-2734/paper6.pdf

[13] Song, M., Günther, C.W., Van Der Aalst, W.M.: Trace clustering in process mining. In: Lecture Notes in Business Information Processing. vol. 17 LNBIP, pp. 109–120 (2009). https://doi.org/10.1007/978-3-642-00328-8_11

[14] Sun, J.C.Y., Lin, C.T., Chou, C.: Applying learning analytics to explore the effects of motivation on online students' reading behavioral patterns. International Review of Research in Open and Distributed Learning **19**, 209–227 (4 2018)

[15] Trauzettel-Klosinski, S., Dietz, K., the IReST Study Group: Standardized Assessment of Reading Performance: The New International Reading Speed Texts IReST. Investigative Ophthalmology & Visual Science **53**(9), 5452–5461 (08 2012). https://doi.org/10.1167/iovs.11-8284

[16] Weijters, A., van der Aalst, W.M.P., de Medeiros;, A.K.A.: Process Mining with the HeuristicsMiner Algorithm (2006)

[17] Xu, J., Liu, J.: A Profile Clustering Based Event Logs Repairing Approach for Process Mining. IEEE Access **7**, 17872–17881 (2019). https://doi.org/10.1109/ACCESS.2019.2894905

[18] Yang, A.C.M., Chen, I.Y.L., Flanagan, B., Ogata, H.: Automatic generation of cloze items for repeated testing to improve reading comprehension. Educational Technology & Society **24**(3), 147–158 (2021)

[19] Yu, C., Balakrishnan, R., Hinckley, K., Moscovich, T., Shi, Y.: Implicit bookmarking: Improving support for revisitation in within-document reading tasks. International Journal of Human-Computer Studies **71**, 303–320 (2013). https://doi.org/10.1016/j.ijhcs.2012.10.012