

Interaction of reading and assessment behavior




Dennis Menze¹, Niels Seidel¹, and Regina Kasakowskij¹

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. In this paper, we examine the interactions between reading and quiz activities 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. Using these session clusters, we further clustered students by their reading and quiz behavior over six time periods within the semester. The results hypothesize a personalization for seven groups of learners characterized by their temporal activity and predominant quiz and reading behavior.

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 [Ko16]) 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 has hardly been investigated so far. We are using a Learning Analytics approach to analyze the interactions between reading and quiz activities. 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.

¹ FernUniversität in Hagen, Chair of Cooperative Systems, Universitätsstr. 1, 58097 Hagen, Germany
{dennis.menze, niels.seidel, regina.kasakowskij}@fernuni-hagen.de,  <https://orcid.org/0000-0003-0002-868X>,  <https://orcid.org/0000-0003-1209-5038>,  <https://orcid.org/0000-0003-2368-9851>

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 first 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.

We examine whether learners evolve in their learning behavior and thus also develop behaviors over the course of a semester that fit their particular life or study situation and their learning progress. To analyze these potential changes in behavior, we consider sequence patterns over six phases. Across these phases, we try to identify learners with similar behaviors using clustering approaches to answer the second research question (RQ2): What changes of reading and quiz patterns can be observed over time? Through this analysis, we aim to find hypotheses for personalized learning opportunities.

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, [BAO19] 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. [SLC18] 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. [Yu13] 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. [Fo14] 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), [Ya21] 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.

Since learning processes in universities typically take place over several months, the question arises as how to analyze changes in learning activities and behaviors over time. [HDH21] analyzed collaborative writing behavior in small groups (N=65). The authors considered learners' individual sequences of defined behavior types (e.g. text contribution, monitoring, inactivity) as vectors. The distance of these vectors was used for clustering participants with similar behavior. However, co-occurring behaviors could not be considered and behavior types were considered as equivalent because of the applied Levenshtein distance measure. In a large study (N=13,362) of 412 courses in China, [Ch17] 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. [BD18] discovered temporal study patterns by coding activities related to video and quiz usage in a MOOC over a period of six weeks. Participants' activities were classified per week and then clustered on sequences of similar learning activities using an agglomerative clustering approach.

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 has 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 [Ha21] and 23 multiple-choice questions were provided, both referred as quizzes.

Data collection & 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. Fig. 1 shows an example of a reading session indicating the beginning and end of sessions in comparison to shorter activity breaks. From these sessions, the events shown in Tab. 1 were derived (10,079 in total).

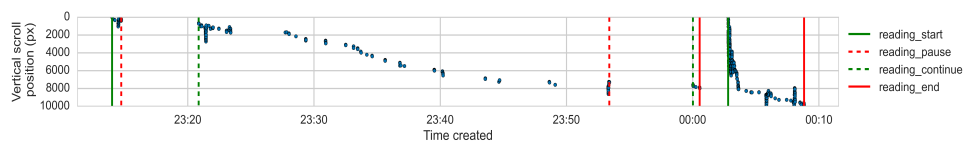


Fig. 1: Reading sessions derived from scroll events indicating begin, end, and activity breaks

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 ([SGV09; XL19], where a trace or profile could be the vector of frequencies of events per session, or the so-called hypothesis-driven approach from [BD18] 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 sequences 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 reading 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 ([Ro87]). The distortion score is defined as the mean of the sum of squared distances of data points to the center of the cluster.

Event	Description	N
reading_start	first scroll event on page after at least 10 min without scrolling	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 last success or fail	313
quiz_success	submitted a solution that is more than 80 % correct (cf. [Ma18])	1,473
quiz_fail	submitted a solution that is less than or equal to 80 % correct	1,350

Tab. 1: Definition of reading and quiz events within a session

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 [WAM06] 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:

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

Since only pairwise activity sequences are considered when looking at state transitions, we also analyzed longer activity sequences using PrefixSpan algorithm [Pe01] 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.

3.3 Transitions over time

In the next step, the classified user sessions were assigned to six different time periods. As we considered due dates of course assignments as a natural temporal structure compared to equidistant periods (e.g. every 2 or 4 weeks) or periods with an equally distributed number of activities, the first four periods (P1–P4) were defined by due dates, whereas the

remainder of the semester P5 and P6 stretched over one month each. For each period, the mode of user session classes was calculated, resulting in a table with 142 users (rows) and 6 periods (columns), each cell containing one of four values (“Mainly quiz”, “Mainly reading”, “Reading and quiz”, “No reading/quiz”). Using one-hot encoding on this table, k-means was used to cluster these dominant user session classes per time period.

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 first part of the section. In the second part, the same activities are considered over time. The anonymized data and analysis scripts used are publicly available at <https://github.com/nise/delfi22>. For a descriptive analysis of the quiz activities, we refer to one of our previous papers [Ha21]. Fig. 2 shows how much of the estimated reading time (1006 characters per minute) 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.

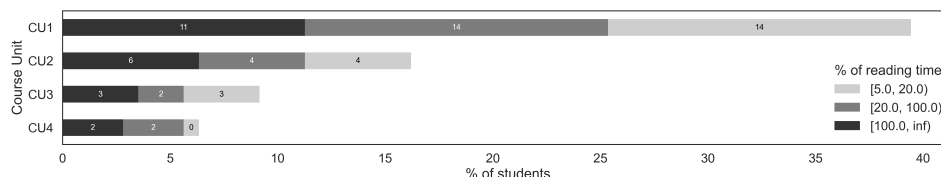


Fig. 2: Percentage of estimated reading time spent on page

4.1 Mining processes and sequences

For k ranging from 2 to 20, we used a silhouette analysis [Ro87] to find the best k representative clusters for the supplied dataset, using measurements of average 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 [WAM06] 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).

SC1 only contains quiz activities. 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.

Session clusters	SC1	SC2	SC3	SC4	SC5	SC6	Total
Sessions (%)	234 (17.2)	207 (15.2)	394 (28.1)	261 (19.2)	169 (12.4)	94 (6.9)	1359 (100)
Users (%)	82 (57.8)	71 (50.0)	95 (66.9)	71 (50.0)	34 (23.9)	48 (33.8)	142 (100)

Tab. 2: Session clusters (SC)

Reading-only activities are part of SC2 (see an example session in Fig. 1). One-fifth each starts reading for a long and medium duration of time. 60 percent 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. 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 terminate 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 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 aim to take multiple quizzes.

SC5 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. 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 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. No further findings could be derived from the eight mined sequences. The user sessions assembled in this session cluster are characterized by reading multiple course units within a session.

4.2 Transitions over time

For $k=7$, the best result was obtained (silhouette score = 0.33, distortion score = 169.31). As depicted in Fig. 3, the size of the clusters C1 to C7 represents the number of learners with a similar learning behavior concerning quiz and reading activities. 9 to 31 learners have been assigned to a cluster. Tab. 3 provides an overview of the distribution of predominant activities in the individual clusters. The session clusters SC1–SC6 described in the previous subsection could be identified in each of the time-related clusters.

Clusters	C1	C2	C3	C4	C5	C6	C7	Total
Learners	9	20	22	28	31	10	22	142
Quiz sessions	106	123	53	441	76	79	11	889
Reading sessions	125	25	47	51	12	13	28	301
Mixed sessions	108	8	20	24	5	0	4	169
Total sessions	339	156	120	516	93	92	43	1359

Tab. 3: Time-related clusters C1 to C7

The clusters are characterized by temporal activity patterns and the predominant activities in the corresponding user sessions. Temporal activity is represented by continuous learning activity for clusters C1 and C4. In contrast, the learners in cluster C5 have been active in hardly more than one period. Learners in clusters C3, C6, and C7 limited their activity to about two periods. The comparatively short activity span is related to an early course drop-out (C3, C5, C6), activity breaks (C2, C3, C5, C7), and a delayed start (C2 and partly C3). Preferences regarding quizzes and reading activities become apparent in Fig. 3. Learners in clusters C4 and C5 show a clear preference for quizzes. In contrast, the 10 learners in C6 started with reading in Moodle in the first two periods, but not all of them continued any activity in the subsequent periods. C1 is the only cluster with a clear preference for a combination of reading and quiz activities, while C2, C3, and C7 also show a considerable mix of activities within one period or over multiple periods. A change in behavior over time becomes partly apparent in C3, where learners move from reading to quizzes. In C1, the learners switched from a quiz to a combination of reading and quiz.

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. 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), 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.

The second part of the analysis was concerned with behavioral changes over time. We decided on one-hot encoding for clustering user sessions over time. A more fine-grained timely similarity of user sessions as proposed by [HDH21] failed due to the large variety of different behaviors. Subsequent clustering of entire user sessions and user sessions over time with different features was caused by a different focus. Session clustering is aimed at answering RQ1 to identify general behavioral patterns about quiz and course text use. Therefore, the selected features represent structural differences in behavior regarding user sessions instead of creating static learner profiles showing a non-existing constant behavior. Since we identified a variety of six different session clusters, we were not able to consider all possible combinations of sessions over time for only 142 participants. The high number of clusters and comparatively small size indicates a limit for clustering over time using data of only 142 participants. As mentioned, the high dimensional data space derived from a few samples limits the significance of the analysis. For this reason, we reduced the number of features and increased the granularity of periods.

From an educational perspective, the results help to explain overall variations in learner participation in a distance learning course. Representative user sessions could be identified through clustering. 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.

Regarding RQ2, the analysis confirms the tendency to study independently of the temporal semester structure intended by the teacher. The students' life, work, and study circumstances require a high degree of temporal flexibility, which is reflected in delayed participation, early drop-out, and short trial periods in the session clusters. Teachers may attempt to further flex their teaching and, for example, avoid deadlines. However, teachers will not be able to accommodate too many different groups due to capacity constraints. Instead, personalized learning prompts and recommendations for learning activities could be created based on the groups and learning partners could be recommended.

The analysis underlines the importance of assessments in terms of quizzes. Even those groups who refused to read course texts in Moodle made use of the quizzes. The participants who refused the quizzes may have submitted other assignments that have been corrected by tutors. In the present analysis, we did not consider semantical relations between quizzes and course texts when clustering.

After quizzes, the provided reading facilities proved to be important for about one-third of the participants. Although the participants tend for short reading phases, considerable medium and long reading spans could be observed. However, some clusters of participants seem to refuse to read course texts in Moodle. Likely, they do not refuse to read at all, since they successfully accomplished some of the quizzes. Instead, they may have used the print or PDF versions of the course texts. Note, that these materials 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 and also over time. 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 included hyperlinks to the corresponding text sections. From the perspective of self-regulated learning, these cross-links are important.

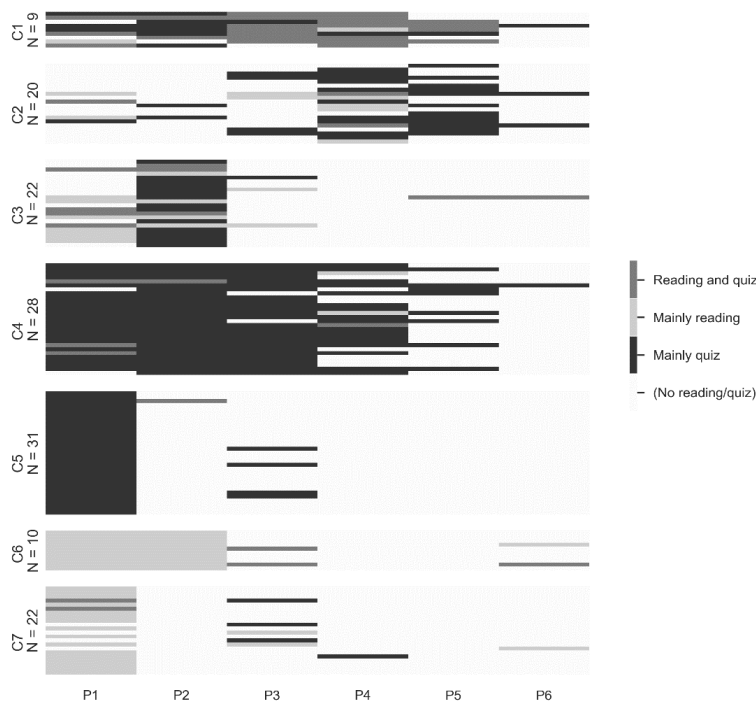


Fig. 3: Clustering users (one row per user) by their quiz and reading activity over six periods

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. Some transitions between reading and quiz patterns could be observed over time (RQ2): early dropouts (C3, C5-C7), later-comers (C2), and a small cluster with some variance (first quiz, then more reading, C1), but the majority behaves like the big cluster (C4) with low variance (mostly quizzes).

In a further study, the found patterns (session clusters and activities over time) should be tried to be replicated in the following semester of the same course with different students. The prediction of clusters over time with machine learning methods, e.g. decision trees, could be elaborated, to investigate possible interventions more concretely. Correlations of found patterns with other factors like grades, assignment results and course re-enrollment should be studied. Furthermore, existing cross-links between quizzes and texts could be investigated. With the presented two step approach, we hypothesized groups of learners with similar reading and quiz behavior that could be provided with a personalized learning environment as well as adaptive support considering individual user sessions. Adaptation could scaffold flexible participation, self-regulation through the close interaction of reading and quiz, and maybe avoid drop-outs using predictions.

Acknowledgements This research was supported by the Research Cluster “Digitalization, Diversity and Lifelong Learning – Consequences for Higher Education” (D²L²) of the FernUniversität in Hagen, Germany.

References

- [BAO19] Boticki, I.; Akcapinar, G.; Ogata, H.: E-book user modelling through learning analytics: the case of learner engagement and reading styles. In: *Interactive Learning Environments* Bd. 27, Taylor & Francis (2019), pp. 754–765
- [BD18] 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*. New York, NY, USA: ACM, 2018, pp.

- [Ch17] 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. In: *Interactive Learning Environments* Bd. 25, Taylor & Francis (2017), pp. 176–188
- [Fo14] 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. In: *Comput. Hum. Behav.* Bd. 41, Elsevier (2014), pp. 478–485
- [Ha21] Haake, J. M.; Seidel, N.; Burchart, M.; Karolyi, H.; Kasakowskij, R.: Accuracy of self-assessments in higher education. In: *DELFI 2021 – Die 19. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.* (2021), pp. 97–108
- [HDH21] Hoppe, H.U.; Doberstein, D.; Hecking, T.: Using Sequence Analysis to Determine the Well-Functioning of Small Groups in Large Online Courses. In: *Int. J. Artif. Intell. Educ.* Bd. 31 (2021), Nr. 4, pp. 680–699
- [Ko16] 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, 2016*, pp. 31–40
- [Ma18] Maldonado-Mahauad, J.; Pérez-Sanagustín, M.; Kizilcec, R.F.; Morales, Nicolás; Munoz-Gama, Jorge: Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. In: *Computers in Human Behavior* Bd. 80 (2018), pp. 179–196
- [Pe01] 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. In: *Proceedings – Int. Conf. on Data Engineering* (2001), pp. 215–224
- [Ro87] Rousseeuw, P.; Rousseeuw, P.J.: Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Comput. Appl. Math.* 20, 53-65. In: *Journal of computational and applied mathematics* Bd. 20 (1987), pp. 53–65
- [SGV09] Song, M.; Günther, C.W.; Van Der Aalst, W.M.P.: Trace clustering in process mining. In: *LNBIP* Bd. 17 LNBIP, 2009, pp. 109–120
- [SLC18] Sun, J. C.Y.; Lin, C.T.; Chou, C.: Applying Learning Analytics to Explore the Effects of Motivation on Online Students’ Reading Behavioral Patterns. In: *IRRODL* Bd. 19, Athabasca Univ Press (2018), pp. 209–227
- [WAM06] Weijters, A. J. M. M.; Aalst, W. M. P. van der; de Medeiros, A. K. A.: Process Mining with the HeuristicsMiner Algorithm, Technische Universiteit Eindhoven, Tech. Rep. WP, vol. 166, 2006, pp. 1–34
- [XL19] Xu, J.; Liu, J.: A Profile Clustering Based Event Logs Repairing Approach for Process Mining. In: *IEEE Access* Bd. 7, IEEE (2019), pp. 17872–17881
- [Ya21] Yang, A. C. M.; Chen, I. Y. L.; Flanagan, B.; Ogata, H.: Automatic Generation of Cloze Items for Repeated Testing to Improve Reading Comprehension. In: *Educational Technology & Society* Bd. 24, IFETS (2021), Nr. 3, pp. 147–158
- [Yu13] Yu, Ch.; Balakrishnan, R.; Hinckley, K.; Moscovich, T.; Shi, Y.: Implicit bookmarking: Improving support for revisitation in within-document reading tasks. In: *International journal of human-computer studies* Bd. 71, Elsevier (2013), pp. 303–320