UNDERSTANDING STUDENT SLIDE READING PATTERNS DURING THE PANDEMIC

Boxuan Ma, Min Lu and Shin'ichi Konomi
Kyushu University
Japan

ABSTRACT

The COVID-19 pandemic has resulted in school closures all across the world, and lots of students have shifted from conventional classrooms to online learning. With the help of ICT technologies nowadays, learning online can be more effective in a number of ways. However, most of the online learning environments without instructors' attention may result in different learning patterns compared to the traditional face-to-face classroom. In this paper, we aimed at detecting the slide reading behaviors of the students by analyzing operational event logs from a digital textbook reader for a lecture offered in our university. We compared reading patterns between traditional face-to-face lectures and hybrid online lectures, our results show that online lectures lead to more off-task behaviors. Our analysis provides a rich understanding of e-book reading and informs design implications for online learning during the pandemic. The findings can also be used to improve the instruction designs and learning strategies.

1. INTRODUCTION

The COVID-19 pandemic has changed education in many aspects. It has resulted in school closures all across the world and students are forced to learn out of the classroom. As a result, education has changed dramatically, with the distinctive rise of e-learning, whereby teaching is undertaken remotely and on digital platforms. Even before COVID-19, there existed substantial infrastructure for online education in many countries. However, no university was ready for a complete shift to online education. Many universities are now delivering course content through various platforms and professors are using online educational platforms, videoconferencing software, and social media to teach their courses (Patricia, 2020; Chakraborty et al, 2020; Bao, 2020). Online educational platforms, like Google Classroom (Iftakhar, 2016) and Moodle (Costa et al, 2012), allow professors to share notes and multimedia resources related to their courses with students. The online educational platforms also allow students to turn in their assignments and professors to keep track of the progress of the students. Videoconferencing tools, like Google Meet, Zoom, and Microsoft Teams, help in organizing online lectures and discussion sessions. Some universities are also disseminating course material through their websites and their own learning management system. While some believe that the unplanned and rapid move to online learning will result in a poor user experience that is unconducive to sustained growth, others believe that a new hybrid model of education will emerge, with significant benefits. The success of MOOCs shows evidence that learning online can be more effective in a number of ways. This is mostly due to the students being able to learn at their own pace, going back and re-reading, skipping, or accelerating through concepts as they choose.

Nevertheless, others argue that homeschooling and online learning are difficult for the students because a structured environment is missing and they are more easily distracted. With regard to traditional face-to-face classrooms, online learning environments that without instructors' attention may lead to more off-task behaviors. Off-task behaviors can be defined as any actions that a student exhibit in the learning environment that is not according to the tasks given by the lecturer. Off-task behavior is a common problem that online classrooms often face. For example, without instructors' attention, devices like computers, mobiles, tablets, etc. can be a reason for distraction because students may play games, use other applications, and browse the

internet. Engaging with off-task behaviors has also been shown to be associated with poor learning (Baker, et al, 2004; Cocea et al, 2009).

Not only different from the traditional face-to-face classroom but also the fully time-free MOOC course, understanding how students learn in such a new hybrid model is still a challenge for researchers. Besides, students are using different study approaches to achieve a specific learning task. Understanding these approaches is important for designing further interventions.

In this paper, we aimed at investigating student reading patterns by using students' reading logs that were collected from a digital textbook reader and compared their reading patterns in traditional face-to-face lectures and hybrid online lectures. Our analysis provides a rich understanding of e-book reading and informs design implications for future e-learning and e-book systems in the age of online learning. The findings can also be used to improve the instruction designs and learning strategies. In addition, by analyzing the different reading patterns of students, interventions can be designed to help off-task students.

2. RELATED WORK

Analyzing learning patterns is important for a better understanding of learners' behaviors and experience, which can help researchers facilitate the design of learning systems, materials, or activities (Sutcliffe et al, 2016). With the increasing use of digital learning materials in higher education, the accumulated operational log data provide opportunities to analyzing how students learn with e-books.

Many researchers focus on studying user reading patterns and their implications from the log data recorded when learners interact with e-book systems to better understand how students read and what they need when reading learning materials. Oi et al. (2015) reported their analysis results of preview and review patterns in undergraduates' e-book logs. Ma et al. (2020) conducted a study to understand the page-flip behavior of students, their results show that students have different page-flip preferences. Shimada et al. (2016) conducted a study based on non-negative matrix factorization and identified five kinds of browsing patterns. Taniguchi et al. (2019) found that highlighters marked by the users on e-textbooks are useful and they provide highlighter recommendations on e-textbooks to help students with better learning. Cheng and Tsai (2014) collected video-recorded data, and used clustering to analyze "Book reading action patterns." Akçapinar et al. (2020) explored students' study approaches and identified three groups of students who have different study approaches, including deep, strategic, and surface. In their study, the relationship between students' reading behaviors and their academic performance is also investigated by using association rule mining analysis. Yin et al. (2019) grouped students were into four clusters using k-means clustering, and their learning behavioral patterns were analyzed.

Based on user behavior analysis, some researchers also find strong correlations between user behaviors and student performance. Brinton et al. (2015) use clickstream data and learner behaviors to predict student performance in a MOOC platform. Okubo et al. (2017) propose a method for predicting final grades of students by a Recurrent Neural Network (RNN) from the log data stored in the educational systems. In addition, Crossley et al. (2016) use click-stream data with NLP tools to predict MOOC completion. Goda et al. (2015) collected data from an e-learning system and used statistical methods to analyze the "Learning pace patterns" of the learners and their relationship to learning outcomes. Seven distinct types of learning behavior were identified, including procrastination, learning habit, random, diminished drive, early bird, chevron, and catch-up. Their results imply that regulated learning could increase learning effectiveness and lead to better learning outcomes.

While existing works provide access to learning log data, they are either aimed at MOOCs or traditional classroom settings, which may differ from the online lecture settings during the pandemic situation. Our analysis focuses on a hybrid online lecture setting, where the instructor gives the course by a videoconferencing tool, and students using an e-book system to read learning materials. Also, we want to compare reading patterns between traditional face-to-face lectures and online lectures.

3. METHOD

3.1 Dataset

The data used in this study were reading logs collected during a 90-minute-long data science course. The instructors and each student used BookRoll (Ogata et al, 2017) (the e-book system in our university) to access teaching slides from a web browser on their personal devices (computer, smartphone, etc.). Other learning activities, such as assignments, quizzes, forum discussions, and so on, were mainly conducted on a Moodle course page. The course lasts for 2 weeks and every lecture has 2 learning materials (teaching slides). Each teaching slide will be used for 45 minutes. In addition, at the end of each lecture, students took part in the quiz related to content. The course was offered to mostly first-year undergraduate university students. In 2019, this course was conducted in a traditional face-to-face classroom. A total of 225 students (mostly freshmen) attended the course in 2019. However, due to the Covid-19 situation, this course was conducted fully online via online meeting software (Zoom) in 2020, the students use BookRoll to access teaching slides at home, keep their microphones muted and cameras off, and listen to the instructor online. A total of 225 students (mostly freshmen) attended the course in 2020. Note that even one is face-to-face and the other is online lecture, the instructors and course contents are all the same in 2019 and 2020.

Figure 1 is the screenshot of BookRoll interface, the basic operation is to flip the page. Students can click the previous button to move to the previous page, and click the next button to move to the subsequent page, they also could use a slider to change pages. Besides that, there are features like bookmark, marker, memo annotating, search, etc. that students can use for learning. All click-stream were recorded in a database that is related to students' interaction with BookRoll. The collected click-stream data contained the following fields. user_id: anonymized student user id. operation_name: the action that was done, e.g. open, close, next, previous, jump, add markers, add bookmarks, etc.). page_no: (the current page where the action was performed), marker: (the reason for the marker added to a page, e.g. important, difficult), role: (the role of the user, including student and teacher), device_code (the type of device used to view BookRoll, e.g. mobile, pc), and operation_date: (the timestamp of when the operation occurred). Table 1 is the description of our dataset.
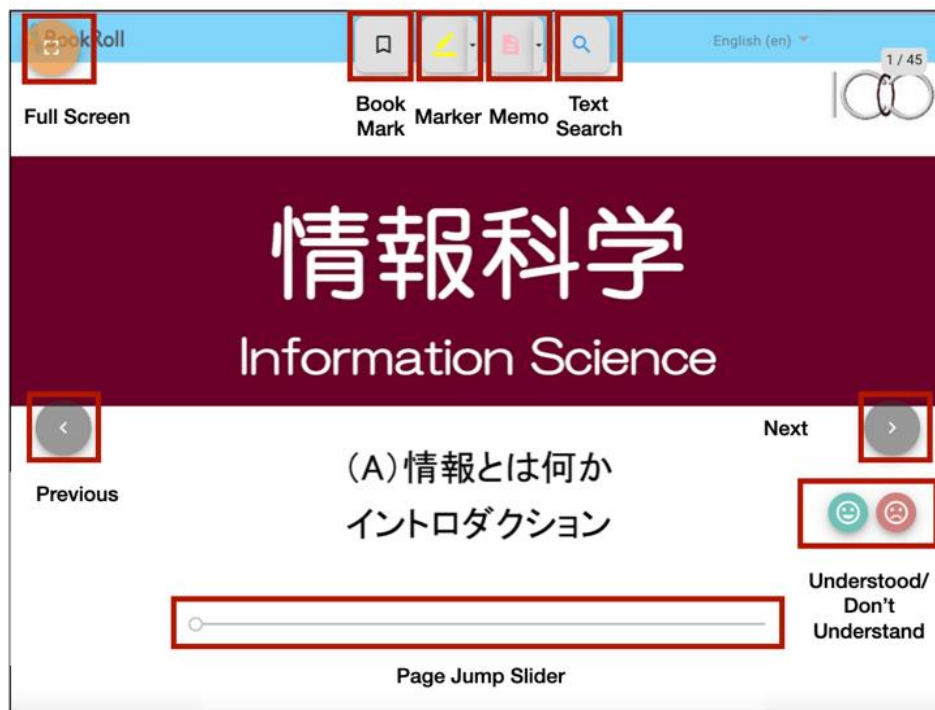

Figure 1. The screenshot of BookRoll interface

Table 1. Description of the Event Stream Dataset

Year	Week	Category	Type	Number
2019	Week 1	Student	Total Student #	175
		Lecture	Lecture Time	90 min
		Operation event	Total Event # (Slide A)	45,141
			Total Event # (Slide B)	31,710
	Week 2	Student	Total Student #	164
		Lecture	Lecture Time	90 min
		Operation event	Total Event # (Slide A)	39,815
			Total Event # (Slide B)	26,846
2020	Week 1	Student	Total Student #	147
		Lecture	Lecture Time	90 min
		Operation event	Total Event # (Slide A)	14,030
			Total Event # (Slide B)	9,656
	Week 2	Student	Total Student #	103
		Lecture	Lecture Time	90 min
		Operation event	Total Event # (Slide A)	12,629
			Total Event # (Slide B)	12,833

3.2 Data Preprocess

For the analysis, we used operation_date (the timestamp of when the operation occurred), page_no (the current page where the action was performed), and role (the role of the user, including student and teacher) columns. First, we grouped the data into 1-minute intervals, then we extracted the pages for all the users, including the instructor and each student. If a user does not have a log for the specific time interval, we assumed that she/he is on the same page where she/he was in the last time. As a result, we got a time-page table for both the instructor and students.

After transforming students' click-stream data into the time-page data, we first calculated all students' reading patterns and relative reading patterns by taking the instructor's reading pattern as a baseline. Next, k-means cluster analysis was used to group students with similar reading patterns. For cluster analysis, we try to group students with similar reading patterns in each teaching slide. The optimal number of clusters is decided based on the elbow method. The details and results will be described in the next section.

4. RESULTS

We visualized all students' reading patterns. Visualizations can be seen in Figure 2. The left is one of the online lectures in 2020 and the right is one of the face-to-face lectures in 2019. The X-axis shows the time, Y-axis shows the page of the books. The intersection of the Time and Page shows the current page of the student at a specific time. Each red line shows the reading patterns of the different students, and the blue line shows the reading patterns of the instructor. As shown in the figure, no matter face-to-face or online lecture, most students' reading pattern is to follow the instructor as we expected, that is, the number of pages increases as time progresses. However, we also observed that there are some students who have different reading patterns.

Enlighted by Akçapınar et al. (2019), we calculated the relative reading patterns of all students to have a better understanding of student reading patterns. To do this we took the instructor's reading pattern as a baseline since the expected reading behavior of students is to follow the instructor during the lecture. The relative reading pattern will be calculated based on the distance of the student's reading pattern and the instructor's reading pattern. For instance, if a student is on page 20 while the instructor is on page 23, then the student's relative distance will be -3. If a student is on page 26, then the relative distance will be +3. And if the student is on page 23 (same page as the instructor), it will be 0.

The results of this calculation are shown in Fig.3. The left is one of the online lectures in 2020 and the right is one of the face-to-face lectures in 2019. The X-axis shows the time of the lecture, while Y-axis shows the student's relative distance from the page where the instructor is currently in. As shown in the figure, both in face-to-face lecture and online lecture, there are some students' reading patterns were ahead of the teacher, some of them tried to follow the teacher, and some were behind of the teacher.

After calculating students' relative distances from the instructor's pattern, we conducted cluster analysis using k-means method to find and compare the common reading patterns in two different learning environments.
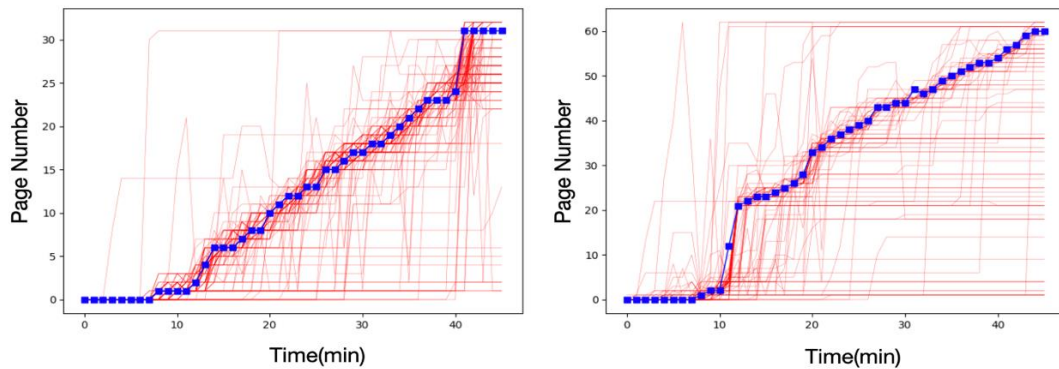

Figure 2. Students' reading patterns across the lecture. Left: 2020 (online lecture). Right: 2019 (face-to-face lecture)
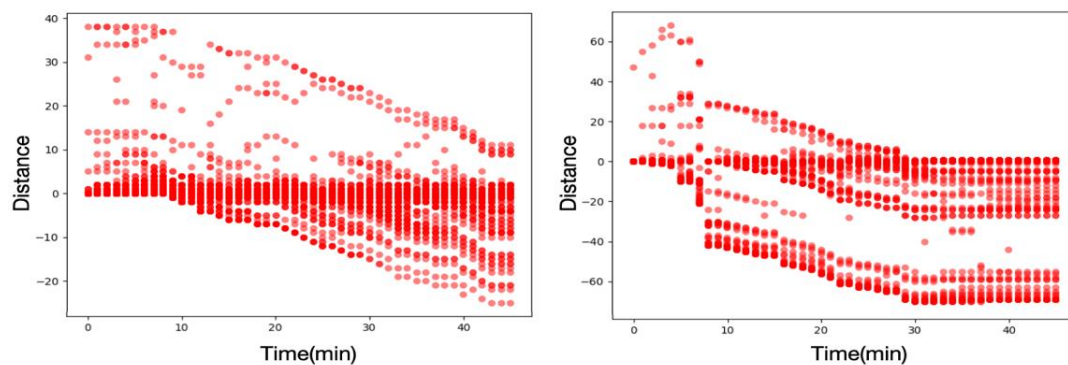

Figure 3. Students' relative reading patterns across the lecture. Left: 2020 (online lecture). Right: 2019 (face-to-face lecture)

4.1 Clustering Result

In order to identify the optimal number of clusters, we have performed the elbow method (Bholowalia et al, 2014) and the number of clusters is set to 4 based on the results. To see the common reading patterns of the students in these clusters, we visualized cluster means as well. Results can be seen in Figure 4. The left is one of the online lectures in 2020 and the right is one of the face-to-face lectures in 2019. The X-axis shows the time of the lecture, while Y-axis shows the cluster means of student's relative distance from the page where the instructor is currently in. From Figure 4, we found four similar patterns both for face-to-face lecture and online lecture. It can be seen that students in cluster 1 (C1) reading the slide even faster than the instructor, while students in cluster 2 (C2) are following the instructor from the beginning until the end of the lecture. In contrast to cluster 2, students in cluster 4 (C4) could be labeled as Off-Task students since they no longer follow the instructor after 15 minutes from the beginning of class. In addition, the relative distance is increased until the end of the course. More interesting thing is that students of cluster 3 (C3), seem to try to catch up with the instructor, and their relative distance is increased first and then decreased.
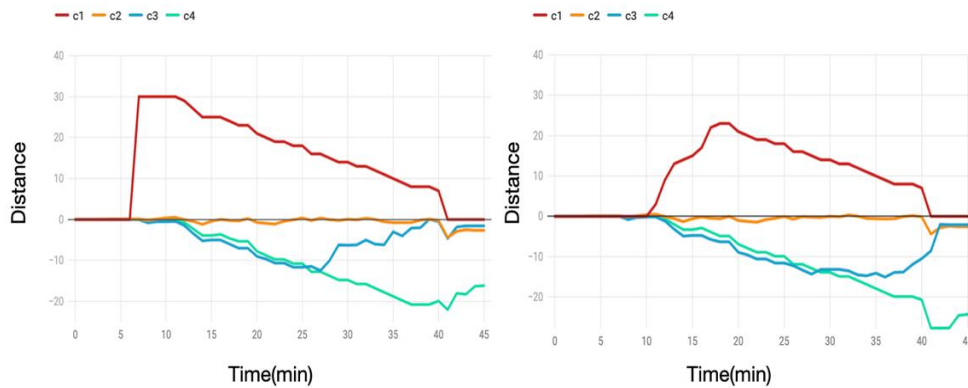
Figure 4. Students' relative reading patterns based on cluster results. Left: 2020 (online lecture). Right: 2019 (face-to-face lecture)

Table 2. Clustering results for 2020 (online lecture)

Category	C1	C2	C3	C4
Week 1 (Slide A)	2%	3%	7%	**88%**
Week 1 (Slide B)	17%	2%	5%	**76%**
Week 2 (Slide A)	14%	7%	3%	**76%**
Week 2 (Slide B)	7%	10%	4%	**79%**

Table 3. Clustering results for 2019 (face-to-face lecture)

Category	C1	C2	C3	C4
Week 1 (Slide A)	20%	31%	5%	**44%**
Week 1 (Slide B)	23%	**43%**	5%	29%
Week 2 (Slide A)	21%	23%	4%	**54%**
Week 2 (Slide B)	26%	24%	14%	**36%**

The percentage of students in each cluster is shown in Table 2 and Table 3. The results in Table 2 highlighted that most of the students had difficulty follow the instructor during the online lectures, more than 76% of students (C4) could not follow the instructor after the first fifteen-minute of the lecture. Only a few students (C2) could follow the instructor all the way during the class. One possible reason could be that a structured environment is missing and they are more easily distracted without the attention of the instructor. The results in Table 3 also show that many students (C4) could not follow the instructor after the first fifteen-minute of face-to-face lectures, however, the percentage of students is lower compared to online lectures. Besides, the percentage of students in C2 is much higher than in online lectures, which hints that students could follow the instructor more easily in face-to-face lectures. Our results show that online format has a significant negative impact on students' persistence in sticking with courses, such an environment that without instructors' attention leads to more off-task behaviors compared to the traditional face-to-face classroom.

4.2 Student Academic Performance

As mentioned before, students took the quiz (full marks=10) at the end of each lecture. We compared the quiz performance of the students in different clusters. Results are shown in Table 4 and Table 5. Since the data were not normally distributed, we used the Mann-Whitney test to compare every two groups. We can see that both for face-to-face lectures and online lectures, students of cluster 2 (C2) get a significantly higher score compared to other clusters, while students of cluster 4 (C4) get a significantly lower score. It is not surprising that students who follow the instructor's reading pattern are more engaging and focusing on the class, and students who give up following the instructor are much easier get distracted during the class and thus receive a low score. However, although leads to more off-task behaviors, our analysis results showed no significant difference in student performance between online and face-to-face students.

Table 4. Clustering results of students' read pattern records for 2020 (online lecture)

Week	Category	C1	C2	C3	C4
Week1	Average Score (SD)	6.09(2.65)	**6.61**(2.55)	5.23(1.81)	5.02(1.97)
Week2	Average Score (SD)	5.16(2.89)	**5.33**(1.39)	4.42(2.58)	3.89(2.68)

Table 5. Clustering results of students' read pattern records for 2019 (face-to-face lecture)

Week	Category	C1	C2	C3	C4
Week1	Average Score (SD)	5.73(1.11)	**5.91**(0.31)	5.51(0.71)	5.18(1.78)
Week2	Average Score (SD)	4.91(1.72)	**5.54**(1.61)	4.45(1.51)	3.91(1.41)

5. CONCLUSION

This research aims to tackle the reading patterns of students while using the e-book system to seek a better understanding of how students read and learn in online lectures and traditional face-to-face lectures. Through the analytics of e-book event stream data, we first calculated all students' reading patterns and relative reading patterns by taking the instructor's reading pattern as a baseline. Next, we conducted a cluster analysis to find the common reading patterns of students. Finally, we compared the quiz performance of the students in different clusters.

The experiment found that both in online and face-to-face lectures there are four reading patterns of students. The results highlighted that online format has a significant negative impact on students' persistence in sticking with courses, such an environment that without instructors' attention leads to more off-task behaviors compared to the traditional face-to-face classroom. Only a few students could follow the instructor all the way during the online lecture. In terms of quiz performance, students follow the instructor received higher score and students show off-task behaviors are associated with poor performance.

Our analysis provides an understanding of e-book reading and informs design implications for learning in the age of online learning. The findings can also be used to improve the instruction designs and learning strategies. In addition, by analyzing the different reading patterns of students, interventions can be designed to help off-task students in online lectures, such as provide early warning settings.

This study has some limitations. First, the sample size is relatively small, which limits the generalizability of the obtained results. Second, while our analysis methods identified e-book reading patterns, we cannot make a strong claim that these clusters are accurately representing students' real intent as reading patterns might depend on other pedagogical methods such as problem sets, discussion, presentation quality, and storyline. Further validation with additional data is required. For future work, we plan to analyze more courses and data streams to analyze the differences in the reading patterns between traditional face-to-face classes and such hybrid online lectures and explore different learning environment's effects on student behavior and learning outcome.

ACKNOWLEDGEMENT

REFERENCES

Akçapinar, G., Chen, M. R. A., Majumdar, R., Flanagan, B., & Ogata, H. (2020, March). Exploring student approaches to learning through sequence analysis of reading logs. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 106-111).

Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Using learning analytics to detect off-task reading behaviors in class. In Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19) (pp. 471-476). Society for Learning Analytics Research (SoLAR).

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: when students "game the system". Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria.

Bao, W. (2020). COVID-19 and online teaching in higher education: A case study of Peking University. Human Behavior and Emerging Technologies, 2(2), 113–115.

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. International Journal of Computer Applications, 105(9).

Boxuan Ma, Jiadong Chen, Chenhao Li, Likun Liu, Min Lu, Yuta Taniguchi and Shin'ichi Konomi (2020). Understanding Jump Back Behaviors in E-book System. Comanion Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK20), pp. 623-631.

Brinton, C. G., & Chiang, M. (2015). MOOC performance prediction via clickstream data and social learning networks. In 2015 IEEE conference on computer communications (INFOCOM) (pp. 2299-2307). IEEE.

Chakraborty, P., Mittal, P., Gupta, M. S., Yadav, S., & Arora, A. Opinion of students on online education during the COVID-19 pandemic. Human Behavior and Emerging Technologies.

Cheng, K. H., & Tsai, C. C. (2014). Children and parents' reading of an augmented reality picture book: Analyses of behavioral patterns and cognitive attainment. Computers & Education, 72, 302 312.doi:10.1016/j.compedu.2013.12.003

Cocea, M., Hershkovitz, A., & Baker, R. S. J. d. (2009). The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? Paper presented at the Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling.

Costa, C., Alvelos, H., & Teixeira, L. (2012). The use of Moodle e-learning platform: a study in a Portuguese University. Procedia Technology, 5, 334-343.

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In Proceedings of the sixth international conference on learning analytics & knowledge (pp. 6-14). ACM.

Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y., & Miyagawa, H. (2015). Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes. Learning and Individual Differences, 37, 72-80.

Iftakhar, S. (2016). Google classroom: what works and how. Journal of Education and Social Sciences, 3(1), 12-18.

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., ... & Hirokawa, S. (2017). Learning analytics for e-book-based educational big data in higher education. In Smart Sensors at the IoT Frontier (pp. 327-350). Springer, Cham.

Oi, M., Okubo, F., Shimada, A., Yin, C., & Ogata, H. (2015). Analysis of preview and review patterns in undergraduates' e-book logs. In 23rd International Conference on Computers in Education, ICCE 2015 (pp. 166-171). Asia-Pacific Society for Computers in Education.

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. In LAK (pp. 598-599).

Patricia, A. (2020). College students' use and acceptance of emergency online learning due to COVID-19. International Journal of Educational Research Open in press.

Shimada, A., Okubo, F., & Ogata, H. (2016). Browsing-Pattern Mining from e-Book Logs with Non-negative Matrix Factorization. In EDM (pp. 636-637).

Sutcliffe, A., & Hart, J. (2016). Analyzing the role of interactivity in user experience. International Journal of Human–Computer Interaction, 33(3), 229–240. doi:10.1080/10447318.2016.1239797

Taniguchi, Y., Shimada, A., Yamada, M., & Konomi, S. I. (2019). Recommending Highlights on Students' E-Textbooks. In Society for Information Technology & Teacher Education International Conference (pp. 874-880). Association for the Advancement of Computing in Education (AACE).

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human Factor Approach. International Journal of Human–Computer Interaction, 35(4-5), 313-322.