Understanding Jump Back Behaviors in E-book System

Boxuan Ma, Jiadong Chen, Chenhao Li,
Likun Liu, Min Lu, Yuta Taniguchi, Shin'ichi Konomi
Kyushu University, Japan
ma.boxuan.611@s.kyushu-u.ac.jp

ABSTRACT: With the increase of digital learning materials in higher education systems, a better understanding of student reading behavior and their effect on student performance get attention. Our research shows that, on average, each e-book system user uses "jump-back" to navigate a course slides for 12.7 times. In this paper, we aim to understand the student's intention for a jump-back. We first formally define the problem of "jump-back" behaviors of reading slide at a face to face lecture, then we systematically study the problem from different perspectives on a real e-Book event stream data. Our study on the dataset reveals several interesting phenomena, e.g. students have different jump-back preferences. Also, students with a higher quiz score were having diverse jump-back behaviors, whereas the students with a comparably low quiz score are feasible to have a comparatively lower jump back frequency.

Keywords: reading behavior; e-Book event stream; educational big data; jump-back.

1 INTRODUCTION

Recently, Learning Management Systems (LMSs) and e-book systems are increasingly used together for supporting daily classroom teaching in many schools. These systems enable us to analyze the log data corresponding to students' learning activities. Such activity log data represent one of the most valuable sources of information for analyzing the activities of students. Analyzing such data provides a novel and great potential for understanding students' behaviors and enhancing education delivery. For example, using clickstream data to predict student performance [Brinton & Chiang, 2015; Okubo et al., 2017], to predict the class completion [Crossley et al., 2016] and clustering learner behaviors [Wang et al., 2016].

Event stream data from e-Book systems have been also utilized to understand students' learning activities. Reading learning materials probably is the most important activity in current college education systems. Actually, the majority of the time that students spend on class is reading slides. Recently, researches have been conducted on the interactions between users and the e-book systems to better understand how students learn and what they need when reading learning materials. For example, pattern mining of preview and review activities [Oi et al., 2015], understanding learning behavior of students [Yin et al., 2015], browsing pattern mining [Shimada, Okubo & Ogata, 2016], and analysis of highlighters on e-textbooks [Taniguchi et al., 2019], etc. However, we found that the jump-back is a frequent behavior with strong user intention. Our preliminary study shows that, on average, each e-book system user uses "jump-back" 12.7 times to navigate a course slide. The reasons may include there is some difficult part that the student cannot understand and the student simply missed some part for other reasons.

In this paper, we conduct a systematic study as a first step to look into this problem in classroom setting by using students' reading logs that were collected from a digital textbook reader in order to better understand student reading behaviors.

The remainder of this paper is organized as follows. In Section 2, we describe the datasets that are used in this paper, then we introduce how we preprocess and analyze the dataset. In Section 3, we conduct the experiments to analyze jump-back behaviors from different perspectives, the details and results also are shown in this section. Finally, we will draw the conclusion and describe future work in Section 4.

2 METHOD

2.1 Data

As the data source, we used reading logs collected from a 90-minutes long in-class activity. Each student used the digital textbook reader (BookRoll) during the lecture. BookRoll is a system that allows digital materials to be delivered in lectures [Ogata et al., 2015; Ogata et al., 2017; Flanagan & Ogata, 2017]. Students can browse anytime and anywhere from a web browser on their personal devices (computer or smartphone). In the BookRoll system, there are features like highlighting, marking, memo function, etc. that students can use for learning. All click-stream were recorded in a database that is related to students' interaction with the system. At the end of the lecture, students took part in the quiz session related to content.

The collected click-stream data contained the following fields: *userid* (anonymized student user id), *contentsid* (the id of the e-book that is being read), *operationname* (the action that was done, e.g. open, close, next, previous, jump, add marker, add bookmark, etc.), *pageno* (the current page where the action was performed), *marker* (the reason for the marker added to a page, e.g. important, difficult), *memo_length* (the length of the memo that was written on the page), *devicecode* (type of device used to view BookRoll, e.g. mobile, pc), and *eventtime* (the timestamp of when the event occurred).

Table 1: Description of the Event Stream Dataset

Category	Type	Number
Lecture	Lecture Time	90 min
	Page Length	83
Student	Total Student #	118
Operation event	Total Event #	263,286
	Total PAGE JUMP #	7,087
	Total SEARCH JUMP #	71
	Total BOOKMARK JUMP #	1,559
	Total NEXT #	154, 401
	Total PREV #	70,360

There are different operations related to our research, i.e. PREV, NEXT, SEARCH JUMP, BOOKMARK JUMP, and PAGE JUMP. PREV means that the student clicked the PREV button to move to the previous page, and NEXT means that the student clicked the next button to move to the subsequent page. Students can also use PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP function to jump to another page. Table 1 lists the statistics of the event stream dataset for a specific lecture. We found that PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP operation is rare in the dataset, instead, students usually click the NEXT or PREV button quickly to jump to the desired page. For example, a student is on page 10 now and he/she wants to jump to page 5, then he would like to click the PREV button 5 times quickly instead of using PAGE JUMP function. To deal with such a problem, we introduced our method in the next section.

2.2 Data Preprocessing

Now we introduce how to preprocess the dataset to deal with the problem above. To start with, we first give the definition of the concept **Complete-jump (Cj)** and other events of which a complete-jump consists.

Definition 1. Complete-jump (Cj). A complete-jump consists of one (or multiple) jump-back actions by a specific student on a specific lecture slide, trying to find the right page to review. Let (s, l, ps, pe) denote a complete-jump, which means student s jumps back from start page ps to end page pe in slide l.

Definition 2. Jumping back (Jb): When a student use PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP function from the current page (ps) to jump back to another page (pe) ($pe < ps$) or click PREV button to go to the previous page, then we say there is a jumping back event.

We noticed that a complete-jump might consist of more than one jump action. For example, the student clicks the PREV button several times to jump back to a previous page. Another example, the student may jump back to a page of no interest and continue to look for the right page that she/he desires to review.

Definition 3. Jumping forward (Jf): When a student use PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP function from the current page (ps) to jump to a page afterward (pe) ($pe > ps$), or click NEXT button to go to the next page, then we say there is a jumping forward event.

There also might be jump-forward actions in a complete-jump. For example, the student jumps back far away from the desired page and then she/he jumps forward to adjust to the right position.

Definition 4. Short watching (Sw): After jumping to the desired page in the slide, the student usually would take a look for seconds. We name it as a short watching event. We use the short watching event to determine the end of a complete-jump. There is a duration period between two jumping events, i.e., from the time the first jumping event ends ($t1$) and the time the next jumping event occurs ($t2$). The duration period should be no longer than Sw, i.e., $t2 - t1 \leq Sw$. In our experiments, we tentatively set $Sw = 2$ seconds[1].

[1] We tried different settings for Sw and empirically selected 2s as an optimal setting.

Actually, complete-jump behavior cannot be obtained straightforwardly. Enlighted by [Zhang et al., 2017], we modified their algorithm based on deterministic finite automaton to reconstruct them.

Based on the definitions above, we use a deterministic finite automaton (DFA) to construct the complete-jump behaviors. Figure 1 shows the state transition in the DFA. There are four states: Ready, Record, Check, Dump. At the Ready state, it stays until receives a jumping back event (Jb), then the state goes to Record. When the state is Record, it maintains a stack. When there are jumping back events (Jb) or jumping forward (Jf) events, it pushes all the events into the stack. If there comes a short watching event (Sw) or some other operations (e.g., the student use MAKER or MEMO function), the state transforms to Check state. When the state is Check, it compares the start page (ps) of the event at the bottom of the stack and the end page (pe) of the event at the top of the stack. If $pe > ps$, the sequence of events in the stack constitutes a jump-forward behavior, then the state goes back to Ready. Otherwise, the state transforms to Dump, where we aggregate the sequence of events in the stack to construct a complete-jump behavior.

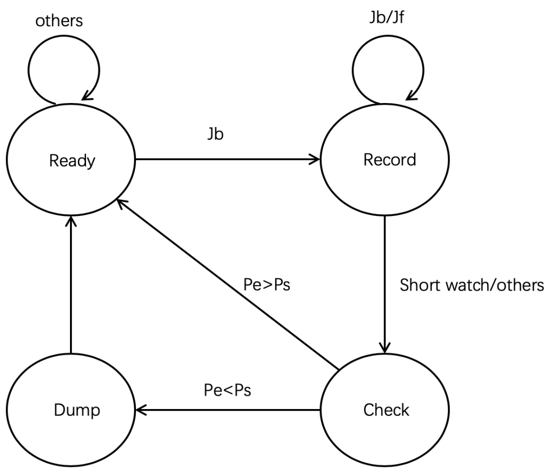


Figure 1: The construction of complete-jump behavior based on DFA

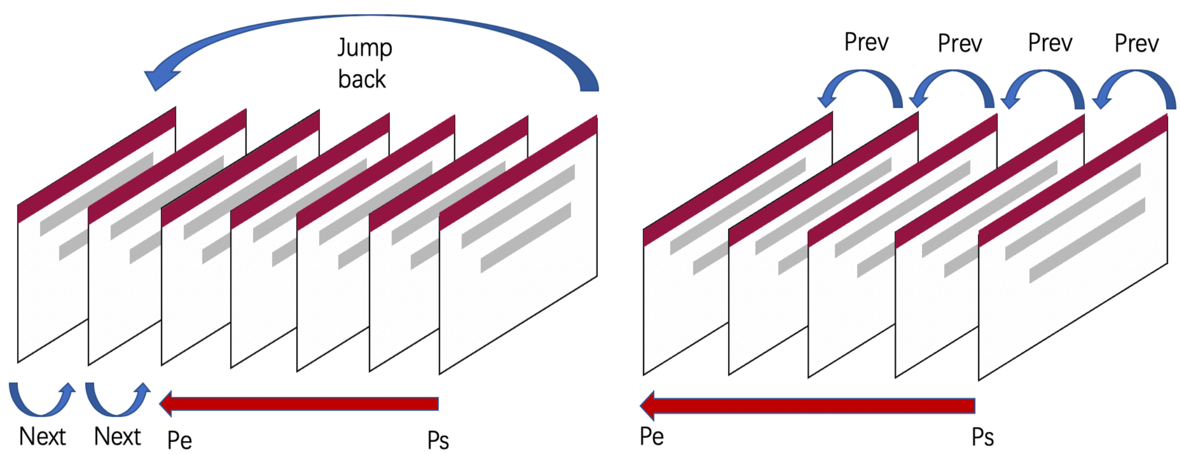


Figure 2: Two complete-jump patterns

Figure 2 shows two common complete-jump patterns in the dataset. The right pattern illustrates a kind of complete-jumps that consist of the event sequence $[Jb, Jb, Jb, Jb]$, which means that the student uses the PREV button 4 times to jump back to a previous page (pe). The left pattern shows a complete-jump that consists of the event sequence $[Jb, Jf, Jf]$. In this kind of scenario, the student uses PAGE JUMP operation to jump back to a previous page firstly and then clicks the NEXT button 2 times to jump to a later page (pe).

2.3 Data Analysis

For the data analysis, first we visualized all students' page flip patterns. Later, we engaged in some investigation of the complete-jump behavior of the students.
The investigations are conducted from three perspectives:

(1) General performances: What is the general performance of students' jump back behaviors? How does the general performance vary in different lectures and slides?

(2) Student preferences: Do students have personal preferences when they jump back? how they show their preferences?

(3) Student Academic Performance: Are there any relationships between students' jump back behaviors and their academic performance?

We analyzed our data by employing basic statistical analysis methods as well as the k-means clustering algorithm to answer the questions above, the details and results will be described in the next section.

3 RESULT

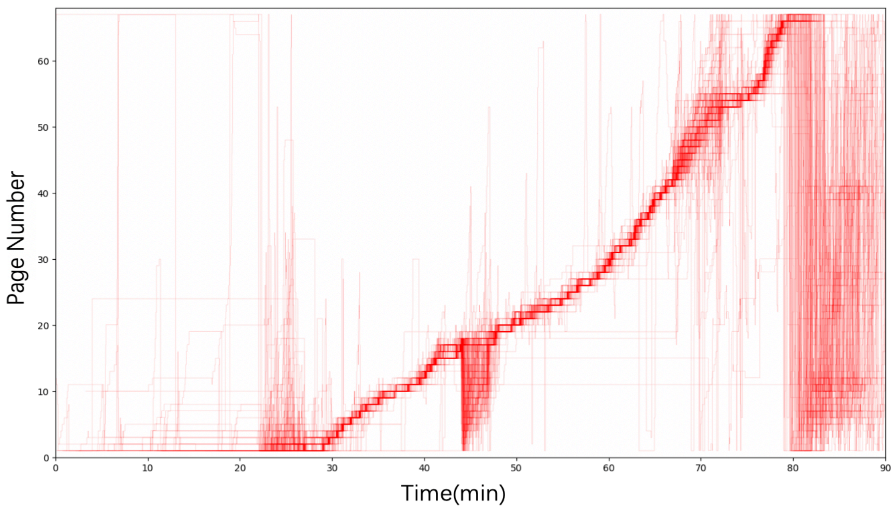


Figure 3: Students' page flip patterns across the lectures

Visualization of page flip patterns of all students can be seen in Figure 3. The X-axis shows the time, Y-axis shows the page of the slides. The intersection of the Time and Page shows the current page of the student at a specific time. Each line shows the reading patterns of a particular student. We can see that many students would like to take a quick look at the entire content in the first 20 minutes of the class, and they review the slide in the last 10 minutes of the class. Figure 3 shows that jump-back is a frequent behavior with strong user intention. Then we will discuss our investigation results of the complete-jump behavior of a student in the following three subsections.

3.1 General Performance

To have a better understanding of students' general jump back performance in a lecture, we plot all the complete-jumps of the slide of a specific lecture in Figure 4. A spot (x, y) represents a complete-jump from the start page x to end page y. The figure shows that most spots are near the diagonal. It indicates that students usually do not jump back to a more distant page from the current page. We name the number of pages between the start page and end page as **jump span**. In this case, the jump span of 80% complete-jumps is smaller than 6 pages, shown as the red area. This phenomenon also exists in other lectures of the dataset.

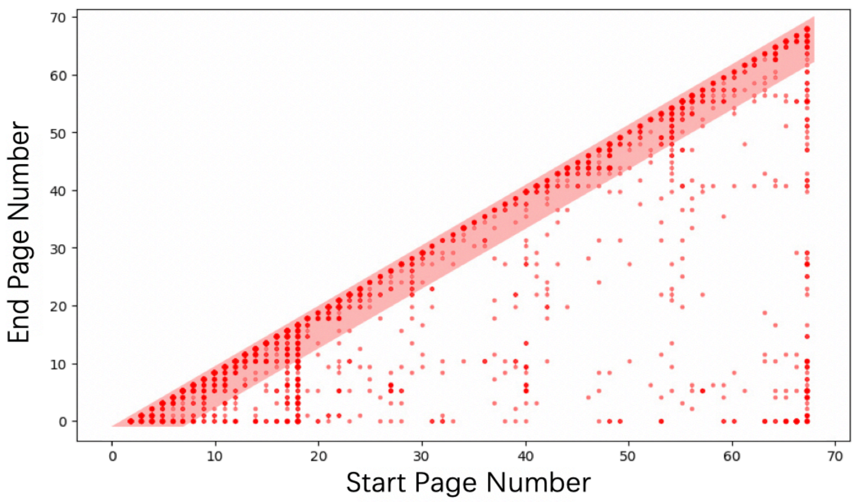
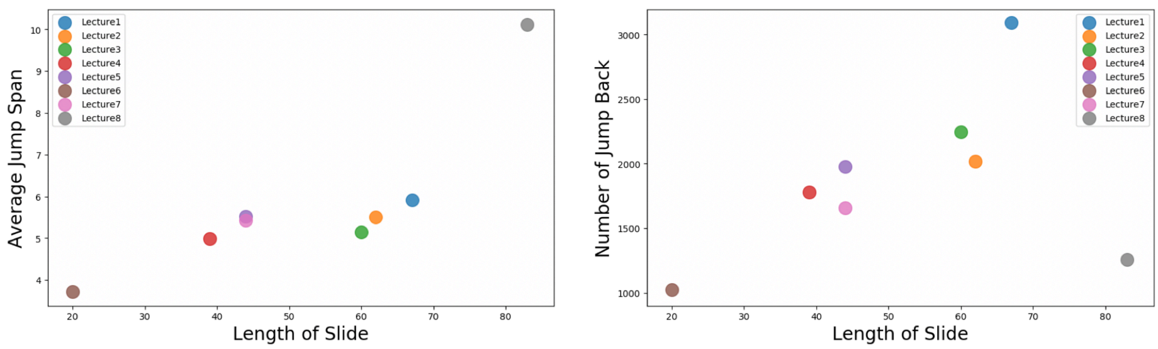


Figure 4: The scatter of complete-jumps. A spot at (x, y) represents a complete-jump from page x to page y.



(a) **(b)**

Figure 5: General complete-jump performance comparison in different lectures. Y-axis: (a) average jump span of each lecture, (b) number of complete-jumps of each lecture. X-axis: the lengths of slides for different lectures.

We have 8 lectures in the dataset. The lengths of slides for different lectures vary from 20 pages to 83 pages. We want to know whether the length of a slide has an effect on complete-jump behavior. Figure 5 (a) shows the correlation between slide length and jump span, and Figure 5 (b) shows the correlation between slide length and the number of complete-jump. The results show that the jump span and the complete-jump number is positively correlated with the length of slides. Where the abnormal occurred in lecture 8 of Figure 5 (b) is conceivable due to the fact that lecture 8 was the last lecture and students need to take the quiz.

3.2 Student Preference

Different students would have different jump-back patterns. For example, impatient students are likely to jump with higher frequency than patient students. To catch students' preferences, we categorize students into different types based on their jump back behaviors leveraging k-means clustering. Table 2 shows the clustering results. In Table 2, Average Stay Time indicates that the average time of reading after the student jumped back to the page.

Table 2: Clustering results of students' jump back records.

Item	C1	C2	C3
# of Jump Back	**18**	8.5	11.6
Max Jump Span	35	28	32
Min Jump Span	1	5	1.1
Average Jump Span	6	**9.7**	7.3
Average Stay Time(s)	39	**292**	101

Students of Clustering 1 have clear preference when they jump back, they prefer to jump more times (18 times) with short jump span (6 pages) and they stay short time after jumping to their desired page (39 seconds), while students of cluster 2 have a preference to jump back farther away (9.7 pages) with lower frequency (8.5 times), but they prefer to stay longer after jumping back to their desired page to have a serious reading (292 seconds). Students of Clustering 3 seem to have no obvious preference and their jump back behavior is more or less unpredictable.

3.3 Student Academic Performance

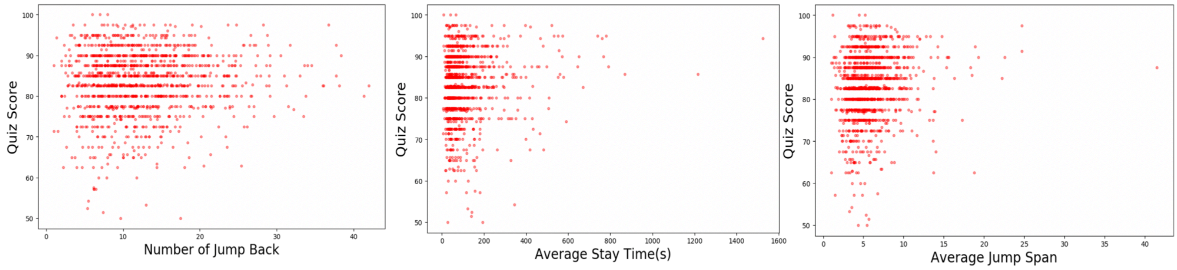
Table 3: Partial correlation results

Item		# of Jump Back	Average Jump Span	Average Stay Time
Quiz Score	PCC	0.1229	0.2422	0.0233
	P-value	0.1848	0.0082	0.2646

As mentioned before, students took the quiz of the last lecture. We use the Pearson correlation coefficient (PCC) to calculate the partial correlation of quiz scores with other variables, such as the number of the jump back and jump span, etc. Table 3 presents the results of the partial correlation analysis. We can see that there is no significant correlation between the jump back behaviors and the quiz score.

Table 4: Comparison results between the high-score group (G1) and low-score group (G2)

Item	# of Jump Back	Average Jump Span	Average Stay Time (s)
G1 Mean (Std)	12.07(6.34)	6.29(3.17)	133.26 (165.58)
G2 Mean (Std)	10.53(5.02)	4.96(2.79)	101.30 (91.15)



(a) **(b)** **(c)**

Figure 6: The scatter of complete-jump behavior and quiz scores. X-axis: (a) the average number of complete-jumps for each student, (b) average stay time for each student, (c) average jump span for each student. Y-axis: the quiz score.

To provide a clear comparison, we plot scatters in Figure 6 of (a) the average number of complete-jumps for each student and quiz score, (b) average stay time for each student and quiz score, (c) average jump span for each student and quiz score. We also split students into the high-score group (G1, score over 90, N=201) and low-score group (G2, score below 70, N=84) to compare the difference of jump back related features in Table 4. Based on the results above, a safe conclusion could be drawn that while the jump back behaviors vary among the students who have a relatively better quiz score, the students with a lower quiz score tend to have a lower frequency of jump back, shorter jump span and stay time.

4 CONCLUSION

This research aims to tackle the reading behavior of learners while using the e-book system to better understand how students read and learn. Particularly, this paper studied the student intention for a jump-back behavior. Through the analytics of e-Book event stream data, we first formally define the problem of "jump-back" behaviors of reading slide at the face to face lecture, then we systematically studied the jump-back behaviors from different perspectives. Our result shows several interesting phenomena, e.g. different students have different jump-back preferences. Students with a higher quiz score were having diverse jump back behaviors, whereas students with a comparably lower quiz score are feasible to have a lower jump back frequency.

As the data source of the current work is limited to page navigation events, it can be difficult to extract a more comprehensive understanding of the jump back behaviors. In our future work, taking other types of the students' behaviors and the teaching processes into account will be helpful to construct explanatory models of irregular page-viewing behaviors. Furthermore, it will invoke new approaches of feedback to the instructors and students, for example, suggestions of useful jump-back destinations of the current page, which will contribute to the improvement of learning efficiency.

REFERENCES

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.

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.

Flanagan, B., & Ogata, H. (2017). Integration of learning analytics research and production systems while protecting privacy. In *The 25th International Conference on Computers in Education, Christchurch, New Zealand* (pp. 333-338).

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.

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.

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).

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

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

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).

Wang, G., Zhang, X., Tang, S., Zheng, H., & Zhao, B. Y. (2016). Unsupervised clickstream clustering for user behavior analysis. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 225-236). ACM.

Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., & Ogata, H. (2015). Identifying and analyzing the learning behaviors of students using e-books. In *23rd International Conference on Computers in Education, ICCE 2015* (pp. 118-120). Asia-Pacific Society for Computers in Education.

Zhang, H., Sun, M., Wang, X., Song, Z., Tang, J., & Sun, J. (2017). Smart jump: Automated navigation suggestion for videos in moocs. In *Proceedings of the 26th international conference on world wide web companion* (pp. 331-339). International World Wide Web Conferences Steering Committee.