Akita J Med 49 : 111-119, 2022 (17)

VISUALIZATION OF VIEWING TRENDS IN HTTP LIVE STREAMING VIDEO MATERIALS FOR MEDICAL EDUCATION USING LOG ANALYSIS

Takumi Kimura, Masayuki Katahira

Department of Medical Information, Akita University Graduate School of Medicine, Akita, Japan

(received 31 March 2022, accepted 01 December 2022)

Abstract

Akita University Graduate School of Medicine always uses HTTP live streaming (HLS) videos for its classes. However, faculty have struggled to understand how students interact with educational streaming video materials. We would like to investigate the possibility to measure viewing trends from accesses confirmed by server operators. Using the HLS system logs, we visualized the viewing trends in this study. From the HLS server, we extracted 165,000 operation logs viewed between April and December of 2016-2018. The logs were then combined for analysis, which resulted in a summary of a series of views for each video and the assignment of 6,409 viewing results. We discovered a positive link between the number of skips and rewinds. The visualization of skipping and rewinding operations on video revealed the concentration of rewinding operations on specific video parts. Clustering of those sequential logs, using video length, viewer rating, skipping, and rewinding as important parameters, revealed three cluster groups ; Label 1 has many "FASTABORT" results (99.5%), Label 3 has many "COMPLETED" results (93.9%), and Label 2 was mixed ; Label 3 dominated at 59.8%. According to our findings, students use current streaming materials moderately. We discuss our findings and limitations to improve future practice (199 words).

Key words : Log analysis, Http live streaming video, Viewing trends, e-learning, Distance learning in medical education

Introduction

Using video materials and e-learning in medical education has become global[1-4]. Class videos distributed and stored in remote classes were regarded as an essential educational resource, and educational streaming distribution services were also been becoming commonplace[1,2]. It was also stated that having students watch video materials before and after class helped them understand the class[5]. Therefore, we presumed that archiving and distributing videos used in remote classes would be a valuable educational tool. The benefits and drawbacks of using asynchronous (pre-recorded) or simultaneous (live broadcasting) video for learning were debated. One educational institution did a survey of scores and questionnaires on simultaneous and asynchronous video streaming classes and discovered no difference in student performance between the two types, implying that the university's choice of video delivery method was not different[6]. It has been claimed that best practices for asynchronous video delivery did not exist, and that faculty had no way of recognizing student interest. However, students preferred asynchronous video because they could watch lectures whenever wher-

Corresponding Author : Takumi Kimura
Department of Medical Information, Akita University Graduate School of Medicine, 1-1-1 Hondo, Akita 010-8543, Japan
Tel : +81 18 801 7021
Fax : +81 18 801 7022
E-mail : t.kimura724@med.akita-u.ac.jp

ever[7-10]. We also investigated the analysis that measured video length and audience ratings, as well as research on videos that allowed learners to skip and rewind[5,11-13]. These studies revealed that the withdrawal rate increased with the length of supplementary video materials in class and that a mechanism was implemented to keep students interested by controlling the teacher's class.

We have been using an e-learning system at Akita University Graduate School of Medicine to provide remote learning for the three core required courses in the doctoral program. Asynchronous video streaming has been used for these contents. Each graduate student could have been required to watch lecture videos remotely and complete a report on the e-learning system form. However, the video viewing evaluation was not confirmed, so we have been researching video viewing trends. In our previous study in 2017, we examined the access logs obtained from the streaming video distribution server in 2016. We reported on the analysis of the environment and period in which students viewed the videos, whether they watched the videos from start to finish, and whether there were any dropouts. After categorizing the viewing trends by the "watch ratio," which was a parameter that could be graphically defined based on the viewing trends of the students, we clarified the relationship between the length of the video and viewing trends[14]. In our previous study in 2018, we contrasted the analysis of viewing trends in 2016 and 2017, as well as "viewing trends" for each class and lecture subject as data acquisition increased[15]. Because our graduate school's video distribution format was asynchronous and did not strictly instruct students on how to view those videos. The goal of this research is to identify the viewing trends that could be used for video tutoring.

Methods

We examined server logs from 2016 to 2018 to conduct additional research on the effects of "video length," and operations like "skip" and"rewind" performed on the video, and the "viewing rate." We also investigated the significance of video manipulation ("skip" and "rewind"). Furthermore, we conducted a comparative analysis to see how "video length" related to the "viewing trends" attached to the logs. Finally, we tried cluster analysis with key parameters ("video length," "watch ratio," "skip," and "rewind") and obtained three cluster labels. The viewing trends label and cluster labels were then cross-analyzed. We properly examined whether it was possible to develop a method for understanding the impact of various parameters in educational video viewing on overall video viewing.

Outline of the Video distribution systems

We installed a streaming server using HTTP live streaming (HLS) protocol committed to class video delivery (Fig. 1)[16]. Using stream-segmented software, this streaming distribution was achieved by sequentially accumulating and transmitting video data that embedded image and audio at 10-second intervals. These intervals were termed transport stream (ts). The index file format playlist file for paging, which specifies the playback order of the divided video files, loaded these ts files in order. The video was retrieved from the delivery server via the http or https protocol and played back.

Operating conditions

The video data collected in each lecture unit was uploaded to the distribution server and linked to the e-learning system, to be accessed from each lecturing unit. There were no guidelines for the length of videos created by teachers. Students could watch the videos

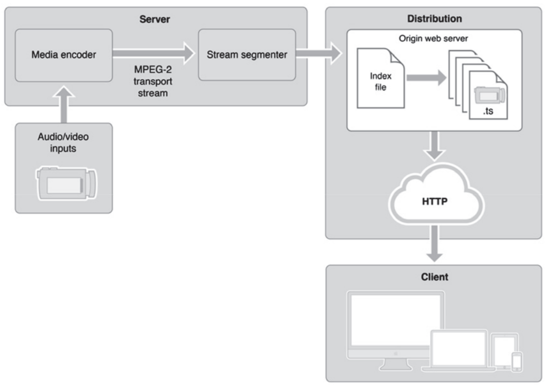

Fig. 1. Outline of the HLS distribution server.

whenever they wanted and wherever they wanted, regardless of the viewing method. The details of the Basic 3 distance learning courses were : Lecture name, (Number of lectures, Lecture title), Latest Medical Research, (16 times, medical01-16), Introduction to Life Science Research, (14 times, life_science01-14), Introduction to Clinical Medical Research, (16 times, clinical01-16). Students sign in to the e-learning system and choose a course. When students select a lecture for each unit, they may be directed to a page that displays a video streaming file in HLS format. Users in either environment could control videos with "start," "skip," "rewind," and "exit" operations.

Log cleansing method and parameters

The research objective was the access logs of the server for HLS video distribution. Individuals could not be recognized from the logs because they were not linked to the logs of the e-learning software. The access protocol for each video on the distribution server was http/https and collected from April to December 2016, 2017, and 2018 (total number of logs approximately n=165,000). The study ran from April to December because it followed the timing of the video release. The logs were cleansed using previous research on HTTP logs and server software logs[5,17]. First, only the connection source, user agent, and referrer fields were extracted individually from an Apache web server combined format. Duplicates were also removed to create a unique list. Before analysis, the user's IP address, hostname, and user agent of the access source in the log were also hidden and treated as unlinked and anonymized. The process was then totally separated from the original access log. Although the referrer field could ascertain which page each video file was accessed from, it was only used in the analysis to establish whether the video tag or video.js was used for viewing. All-access points were URIs on the university's e-learning system's links. We only extracted the video data and index playlist files access. Then, we merged the lecture name and segmented the videos file sequence number from the file path names. Table 1 shows the parameters used in the logs. Fig. 2(a) depicts the format of the combined log file. For analysis to investigate the visualization of video skipping and rewinding operations, logs following the order of accesses without integrating file sequence numbers were in Fig. 2(b).

Table 1. Description of each parameter. The ts files in the table is an abbreviation for transport stream.

Parameter	Description
identifier	ID randomly generated and in sequential order.
date	Date of connection (YYYY/mm/dd).
year	Year of connection (YYYY).
month	Month of connection (mm).
dayofweek	Weekday of connection (Sunday-Saturday).
hour	Connection Time (HH).
title	Connected lecture name.
player	The html video tag used for the connection (JS or video).
firstSeq	First connected ts split file.
maxSeq	Last connected ts split file.
watchedSeq	Number of ts split files viewed.
videoLength	Maximum number of ts files the "title" parameter had.
viewingTrends	Viewing trends keywords arbitrarily set in previous studies ex.)SKIPPED.
watchratio	Percentage of ts files viewed in "videoLength" for each lecture.
skip	Number of times to skip a video.
rewind	Number of times to rewind a video.
skiplist	A record of a sequence of ts files skipped from the beginning of the video.
rewlist	A record of a sequence of ts files rewound from the time of video playback.
Seq	Value of ts (10 sec) in a video.
deltaSeq	Value of difference from previous "Seq".

Log analysis

First, we referred to previous research on how to compare viewing patterns along these time series and consideration of dropout rates[11,17-19]. We evaluated the relationship between the number of skips and rewinds for all accesses, and we visualized the skipping and rewinding operations along the video timeline using the 2017 instructional video "clinical01" as an example. Also, longer video lengths result in higher withdrawal rates[11,20]. So, this was discovered that video length affected viewing trends. To better comprehend the viewing trends, we

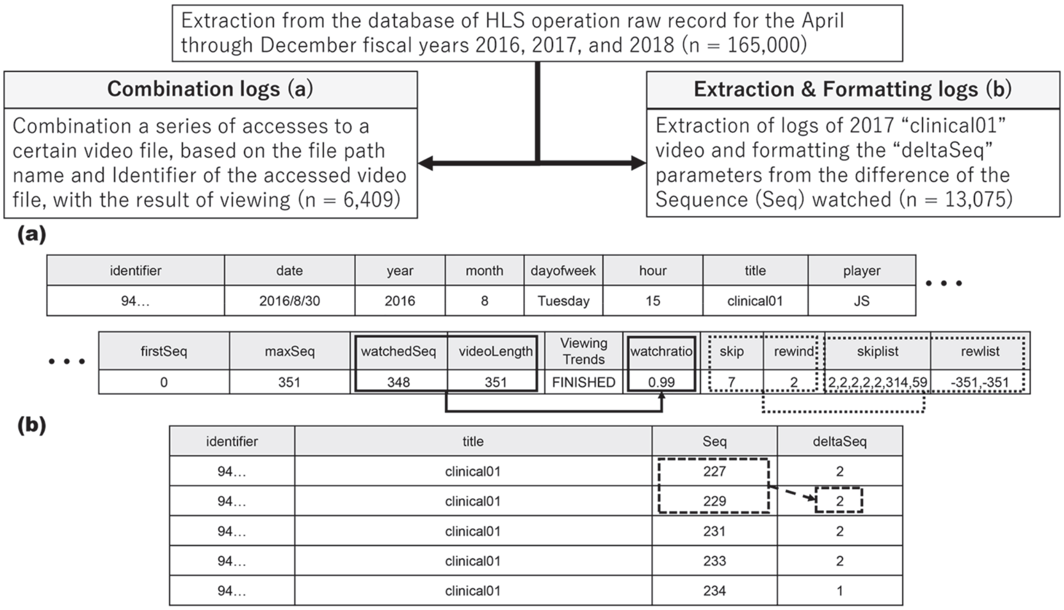

Fig. 2. Log (a) was tied to a randomly assigned identifier and had the date and time of the connection and a sequence of combined views. The "watch ratio" was a value derived from "watched Seq" and "video length." The number of "skiplist" and "rewlist" times was measured by the parameter "skip" and "rewind," respectively. Log (b) follows the sequential number of "Seq" connected to the video without being combined as in (a), and the "delta-Seq" parameter measures the difference between the previous and next "Seq" to indicate how much skipping and rewinding time that occurred.

developed a "watch ratio," a parameter that indicates the percentage of viewing, by performing mathematical processing between parameters[14,15]. The "Watch ratio" was a marker that could reveal what percentage of the video was viewed by dividing the video length of each class video by the number of "ts (the number of separate files of 10 seconds each)" that were viewed[14]. Consequently, we developed keywords, which were the "viewing trends label" of the videos, "COMPLETED," where all split ts files were accessed, "FINISHED," where higher than 98% of the ts files were accessed, "SKIPPED," where the last ts file was accessed, and the access percentage was less than 98%, "ABORTED," where the last ts file was not accessed, and the access rate was less than 98%, and aborted and "FASTABORT," were only less than 20% of the files were viewed[15]. Based on these parameters, we compared the percentage of browsing types. We used a nonparametric Kruskal–Wallis test with a significance level of 0.05 to compare video length between video types. Since video length, number of skips, number of rewinds, and viewing ratio influence video viewing among the parameters we have used, we evaluated if it was feasible to automate the division of video viewing groups through two-step cluster analysis using the AIC (Akaike's Information Criterion). Log data cleansing and data table manipulation were conducted on a Unix (Free BSD) computer server, and analysis was done using IBM SPSS statistics (24.0.0).

Results

In the relationship between the number of "skip" and "rewind" operations in each video viewing log, we discovered a positive correlation. Many "skip" operations were preceded by "rewind" operations (Fig. 3).

To simulate the skipping and rewinding operations, we exempted the logs viewed in sequence (deltaSeq=1) and depicted each operation between "$-60<=$deltaSeq" and

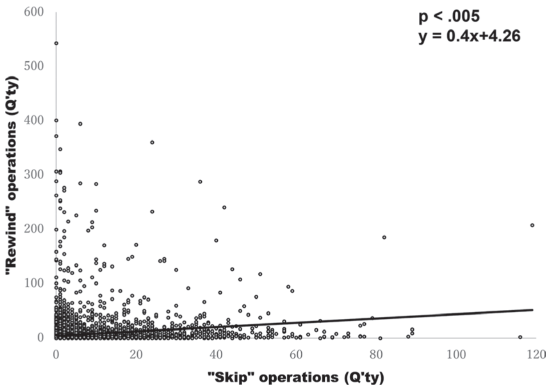

Fig. 3. Numbers of "skip" used for single video access versus numbers of "rewind."

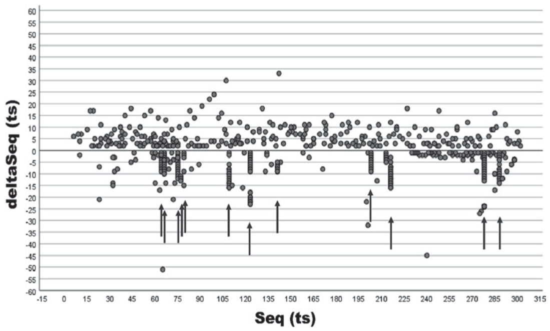

Fig. 4. Using the video "clinical 01" from 2017, we illustrated the skipping and rewinding operations along the video's timeline. At the exact location, the "Seq (ts)" area with arrows could be observed rewinding operation. When tuples with deltaseq=1 that were not skipped or rewound were excluded, the total number of plots was 672.

"deltaSeq<=60" as a scatterplot. In the video of "clinical01" in 2017, we found that the "rewind" operation was recreated in the ts section with arrows. The number of seconds of "rewind" was also adjusted by gradually changing it (Fig. 4).

The video length of viewing in each result was compared using a pairwise nonparametric test. COMPLETED-ABORTED (p<.001), COMPLETED-FASTABORT (p<.001), COMPLETED-FINISHED (p<.001), and COMPLETED-SKIPPED (p<.001) (Fig. 5).

We tried a two-step cluster analysis with the AIC criteria using automatic clustering, the AIC is 9790.718 at cluster number 3 with "videoLength," "watch ratio," "skip," and "rewind" as significant parameters for analyzing the viewing trends. The findings were classified into three cluster labels. We conducted results and cross-tabulations for each label to examine the fitness. We discovered that Label 1 has numerous "FASTABORT" results (1,202 lines), and Label 3 has numerous "COMPLETED" results (2,916 lines).

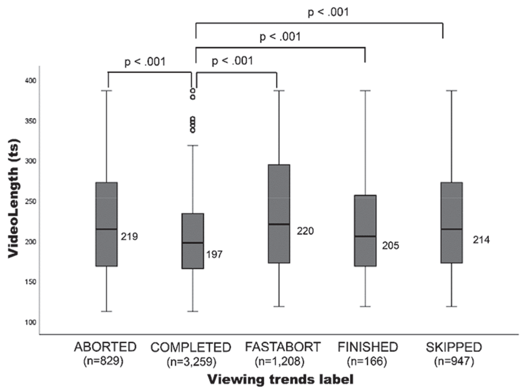

Fig. 5. Comparison of each "result" keyword and "video length." The video length of viewing in each result was compared using a pairwise nonparametric test. COMPLETED-ABORTED (p<.001), COMPLETED-FASTABORT (p<.001), COMPLETED-FINISHED (p<.001), and COMPLETED-SKIPPED (p<.001).

Discussion

In this study, we discovered that the more skips there were, the more rewinds there were, implying that rewinds related to adjustments occurred when skips were used. Visualizing analysis on one video revealed that the rewind operation was performed at a specific time, whereas the skip operation was identified as random. We assumed this was due to reviewing operations at difficult-to-understand points and reviewing due to the need to write a report on the e-learning system. The rewind operation was repeated several times at an appropriate position (ts) after determining the number of seconds (ts) which viewer probably wanted to review. In terms of the maximum video length in the keyword, there was a difference between "COMPLETED" and

Table 2. Results of applying AIC standard cluster analysis with "videoLength," "watchratio," "skip," and "rewind" as elements, representative values were described above the table. At the bottom of the table, cross-tabulation values showing fit with "viewing trends" were described.

Cluster Label	Label 1 n=1,908 (29.8%)	Label 2 n=666 (10.4%)	Label 3 n=3,835 (59.8%)
videoLengthAverage (ts)	215.09	235.50	205.41
watchratio (%)	15	78	96
skip (Frequency)	1.28	28.55	1.86
rewind (Frequency)	3.83	42.77	3.21
FASTABORT	1,202 (99.5%)	6 (0.5%)	0 (0%)
ABORTED	481 (58.02%)	101 (12.18%)	247 (29.79%)
SKIPPED	119 (12.57%)	312 (32.95%)	516 (54.49%)
FINISHED	0 (0%)	10 (6.02%)	156 (93.98%)
COMPLETED	106 (3.25%)	237 (7.27%)	2,916 (89.48%)

other keywords. It was hypothesized that the shorter the period, the lower the dropout rate. As anticipated, there was a marker that the shorter the video length, the lower the decline rate. This was thought to be because of the viewing method. For example, it was likely that the viewer reviewed the video while completing the report, or that the speaker spoke too quickly, causing an important video segment (ts) to be rewound several times.

We obtained three cluster labels through a two-step cluster analysis using"videoLength," "watchratio," "skip," and "rewind" as important parameters for analyzing the viewing trends. We classified "COMPLETED," "FASTABORT," and "FINISHED." "SKIPPED" and "ABORTED" as "viewing trends." Label 1 contains tons of "FASTABORT" and "ABORTED," and was characterized by an intangible amount of skips and rewinds. This clearly represented early viewing dropouts. Label 3 had a majority of "FINISHED" and "COMPLETED," and indicated few skip and rewind signs like Label 1. We obtained a 59.8% ratio that accounts for all viewing sequences, and it was clear that most viewers watched the video without performing any operations. Label 2 did not have the highest proportion of "SKIPPED," but the representative values of skips and rewinds were the highest among labels. Although judging criteria were inevitably biased, we believed that by creating clusters with four factors, we could discern between "COMPLETED" and "FINISHED," and fast escape viewing, such as "FASTABORT.

It might be necessary to analyze viewing trends to determine how videos were used for learning because best practices for video delivery had not been established[7]. Since it has been affirmed that students view those videos numerous times (approximately 1.9 times) to grasp the content, evaluating the student's viewing based on a single viewing result might not be appropriate[21,22]. Furthermore, there were numerous variations in how e-learning video materials were used, with some videos being viewed entirely and others used only to review the slides in the video[23-25]. These were most likely the reasons why faculty members could not consider best practices for video delivery. As a result, it may be necessary to examine viewing trends and review the content, particularly understanding "viewing trends" with analytics will impact how videos are taken to improve student satisfaction.

Limitations

The current work was a case study that attempted to visualize video material viewing trends at Akita University Graduate School of Medicine. Therefore, the findings of our study may not apply to other streaming video log analyses used for medical education. Because the logs could not differentiate between video viewers, the same viewers who had watched the same videos multiple times could not be dismissed. Extending our investigation from previous research, we gathered and analyzed data from 2016 to 2018. We did not use random sampling or other methods because we needed to compile a series of consecutive logs as a series of views. Two-step clusters were chosen for this study. The reason was that we did not use hierarchical clusters because we used the total number of logs (165,000 logs were compiled per sequence, yielding 6,409 logs for clustering), and we chose the four parameters that would influence both viewers and videos for the first time based on our previous video logs studies. The AIC (9,790.718) value was high because the clusters were modeled without establishing the number of clusters. In future research, we would like to investigate a method for determining parameters for optimal modeling.

Conclusions

In this study, we discovered three significant viewing trends based on different important viewing parameters like sequential logs, video length, viewer rating, skipping, and rewinding. Label 1 has many "FASTABORT" results (99.5%), Label 3 has many "COMPLETED" results (93.9%) and Label 2 was mixed ; 59.8% was dominated by Label 3. Hence, our findings indicate that current streaming materials are modestly utilized by students.

Conflict of Interest

The authors have no conflicts of interest to declare.

Acknowledgements

We would like to thank Enago (www.enago.jp) for the English language review.

References

1) Katahira, M. and Nakamura, A. (2010) Statistical evaluations on freshmen's skills and knowledge of ICT and effectiveness of our Class. *Comput. Educat*, **29**, 86-91.
2) Takeda, N., Takeuchi, I. and Haruna, M. (2007) Assessment of learning activities using streaming video for laboratory practice education : Aiming for development of e-learning system that promote self-learning. *J. Pharm. Soc. Jpn.*, **127**(12), 2097-2103.
3) Botelho, M.G., Gao, X. and Jagannathan, N. (2019) A qualitative analysis of students' perceptions of videos to support learning in a psychomotor skills course. *Eur. J. Dent. Educ.*, **23**(1), 20-27.
4) Sowan A.K. (2014) Multimedia applications in nursing curriculum : The process of producing streaming videos for medication administration skills. *Int. J. Med. Inform.*, **83**(7), 529-535.
5) Dev, P., Rindfleisch, T.C., Kush, S.J. and Stringer, J.R. (2000) An analysis of technology usage for streaming digital video in support of a preclinical curriculum. *Proc. AMIA Symp.*, 180-184.
6) Moridani M. (2007) Asynchronous video streaming vs. synchronous videoconferencing for teaching a pharmacogenetic pharmacotherapy course. *Am. J. Pharm. Educ.*, **71**(1), 16.
7) Berner, E.S. and Adams, B. (2004) Added value of video compared to audio lectures for distance learning. *Int. J. Med. Inform.*, **73**(2), 189-193.
8) Bennett, P. and Glover, P. (2008) Video streaming : Implementation and evaluation in an undergraduate nursing program. *Nurse Educ. Today*, **28**(2), 253-258.
9) Liu, T. and Choudary, C. (2006) Content-adaptive wireless streaming of instructional videos. *Multimed. Tools Appl.*, **28**, 157-171.
10) Wynter, L., Burgess, A., Kalman, E., Heron, J. and Bleasel, J. (2019) Medical students : what educational resources are they using ?. *BMC Med. Educ.*, **19**, 36.
11) Hildebrand, J.D. and Ahn, B. (2018) Student video viewing habits in an online mechanics of Materials Engineering Course. *Int. J. Eng. Pedagog.*, **8**(3), 40-59.
12) Alpert, F. and Hodkinson, C. (2018) Video use in lecture classes : Current practices, student perceptions and preferences. *AEM Educ. Train.*, **61**(1), 31-45.
13) Scagnoli, N., Choo, J. and Tian, J. (2019) Students' insights on the use of video lectures in online classes. *Br. J. Educ. Technol.*, **50**(1), 399-414.
14) Katahira, M. (2017) Viewing trends analysis on streaming video materials in e-Learning. *Proc. of 2017 PC Conference*, 240-243.

15) Katahira, M. (2018) Viewing trends analysis and annual comparison of e-Learning streaming video materials. *Proc. of 2018 PC Conference*, 283-286.

16) Katahira, M. (2016) A practical report on linking an e-learning system and a streaming video distribution server using OSS. *Proc. of 2016 PC Conference*, 175-176.

17) McNulty, J.A., Hoyt, A., Gruener, G., Chandrasekhar, A., Espiritu, B., Price, R., Jr and Naheedy, R. (2009) An analysis of lecture video utilization in undergraduate medical education : associations with performance in the courses. *BMC Med. Educ.*, **9**(1), 6.

18) Traphagan, T., Kucsera, J.V. and Kishi, K. (2010) Impact of class lecture webcasting on attendance and learning. *Educ. Technol. Res. Dev.*, **58**(1), 19-37.

19) Beatty, B.J., Merchant, Z. and Albert, M. (2019) Analysis of student use of video in a flipped classroom. *TechTrends,* **63** (4), 376-385.

20) Yudko, E., Hirokawa, R. and Chi, R. (2008) Attitudes, beliefs, and attendance in a hybrid course. *Comput. Educ.,* **50**(4), 1217-1227.

21) Toye, F., Jenkins, S., Seers, K. and Barker, K. (2015) Exploring the value of qualitative research films in clinical education. *BMC Med. Educ.*, **15**, 214.

22) Coyne, E., Frommolt, V., Rands, H., Kain, V. and Mitchell, M. (2018) Simulation videos presented in a blended learning platform to improve Australian nursing students' knowledge of family assessment. *Nurse Educ. Today,* **66**, 96-102.

23) Huh, D., Kim, J.H. and Jo, I.H. (2019) A novel method to monitoring changes in cognitive load in video-based learning. *J. Comput. Assist. Learn.*, **35**(6), 721-730.

24) Li, L.Y. and Tsai, C.C. (2017) Accessing online learning material : Quantitative behavior patterns and their effects on motivation and learning performance. *Comput. Educ.*, **114**, 286–297.

25) Jeske, D., Backhaus, J. and Stamov, R.C. (2014) Self-regulation during e-learning : Using behavioural evidence from navigation log files. *J. Comput. Assist. Learn.*, **30**(3), 272-284.

Abbreviations and Technical Terms

Technical Keywords	Description
HTTP, http or https	Abbreviation of hypertext transfer protocol which is the connection method used to connect to the web site. The https has a secure features.
HTTP Live Streaming (HLS)	Synchronous streaming video delivery with HTTP connections.
.ts (ts)	A standard format for distributing and storing video, audio, and files, known as a transport stream file.
video tag, Video. js	Embedding tags used for embedding videos in web texts.
Apache Web server	WWW server software that runs on servers called middleware, created by the open community, the latest version is available at https://httpd.apache.org/.
URI	Abbreviation of uniform resource identifier, that includes the connection protocol and query string of the connection source in the URL.
Log parameters : connection source, user agent, refferer, IP address, Hostname	Apache software log parameters : Information on the connection source, OS and browser information, Pre-connected URI, connection source information given to network devices and Host name from which to connect.