Li, L.-Y., & Huang, W.-L. (2023). Effects of Undergraduate Student Reviewers' Ability on Comments Provided, Reviewing Behavior, and Performance in an Online Video Peer Assessment Activity. *Educational Technology & Society*, *26* (2), 76-93. https://doi.org/10.30191/ETS.202304_26(2).0006

Effects of Undergraduate Student Reviewers' Ability on Comments Provided, Reviewing Behavior, and Performance in an Online Video Peer Assessment Activity

Liang-Yi Li[1*] and Wen-Lung Huang[2]

[1]Program of Learning Sciences, Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan // [2]Department of Communication, Fo Guang University, Yilan, Taiwan // lihenry12345@ntnu.edu.tw // wlhuang@mail.fgu.edu.tw

[*]Corresponding author

(Submitted January 28, 2022; Revised August 16, 2022; Accepted Aug 29, 2022)

ABSTRACT: With the increasing bandwidth, videos have been gradually used as submissions for online peer assessment activities. However, their transient nature imposes a high cognitive load on students, particularly low-ability students. Therefore, reviewers' ability is a key factor that may affect the reviewing process and performance in an online video peer assessment activity. This study examined how reviewers' ability affected the comments they provided and their reviewing behaviors and performance. Thirty-eight first-year undergraduate students participated in an online video peer assessment activity for 3 weeks. This study analyzed data collected from the teacher's and peer reviewers' ratings, comments provided by peer reviewers, and system logs. Several findings are significant. First, low-ability reviewers preferred to rate higher scores than high-ability reviewers did. Second, low-ability reviewers had higher review errors than high-ability reviewers. Third, high-ability reviewers provided more high-level comments, while low-ability reviewers provided more low-level comments. Finally, low- and high-ability reviewers showed different behavior patterns when reviewing peers' videos. In particular, low-ability reviewers invested more time and effort in understanding video content, while high-ability reviewers invested more time and effort in detecting and diagnosing problems. These findings are discussed, and several suggestions for improving the instructional and system design of online video peer assessment activities are provided.

Keywords: Video peer assessment, Learning analytics, Comments provided, Behavior pattern

1. Introduction

Peer assessment (PA) is a process whereby students assign grades to peers' submissions and provide comments for peers to improve their work (Tenorio et al., 2016). PA can reduce the teacher's workload and improve students' attitudes, critical thinking, and judgement skills (Tenorio et al., 2016), and has already been applied in many disciplines such as science, language, and programming.

With the rapid development of information and communication technologies (ICT), web-based peer assessment approaches have been widely used (Formanek et al., 2017; Hsia et al., 2016a; Hsu et al., 2018). They can help teachers to share the assessment tasks and results and monitor students' progress (Lin et al., 2001). Students can also conduct peer assessment activities on the Web without the limitations of time and space.

Generally, written text is the primary target assessed in online peer assessment activities (Tenorio et al., 2016). However, with the increasing bandwidth of the internet, videos have been gradually used as submissions. In contrast to static text and images, videos provide several advantages for peer assessment activities. First, videos, which present dynamic visual and verbal information, are especially useful for evaluating actions and voices (Hsia et al., 2016b; Lai et al., 2020). Second, video playing interfaces provide multiple operations (e.g., resume, pause, fast forward, and back). Students' operations can be recorded in system logs when reviewing peers' videos. These logs can then be analyzed in order to understand how students review peers' videos (Li, 2019).

Although online video peer assessment provides these advantages, not all students can benefit from it because reviewing peers' assignments on the Web is a self-regulated process in which learners freely control their reviewing path and pace. Students with different individual characteristics may have different behavior patterns when reviewing peers' assignments. These different behavior patterns may also result in different outcomes (Shirvani Boroujeni & Dillenbourg, 2019). To support students with different individual characteristics, teachers should understand how students perform during the Web learning process and what the relationships are between students' individual characteristics and learning behaviors and performance.

Reviewers' ability, which is defined as task-related knowledge and skills, is one individual difference that is often examined in peer assessment activities. Previous studies have shown that reviewers' ability affects the quantity and quality of comments they provide for peers' submissions (Huisman et al., 2018; Patchan et al., 2013). However, these studies primarily examined written text (e.g., writing compositions). Because videos are transient media in which the information presented changes dynamically, students easily experience the problems of cognitive overload and disorientation when viewing peers' videos. Therefore, whether the same effects also happen when the submissions are videos is unknown. In addition, previous studies have found that individual differences such as prior knowledge and cognitive style can affect learners' video watching behaviors and performance when viewing instructional videos (de Boer et al., 2016; de Boer et al., 2011; Li, 2019). Whether students' ability also affects their reviewing behaviors and performance is also unknown. Therefore, the purpose of this study was to examine how reviewers' abilities affect the comments they provide and their reviewing behaviors (e.g., behavior patterns) and performance (e.g., review error and the quantity and quality of the comments provided).

2. Related works

This section first reviews previous studies related to the use of videos in online peer-assessment activities. The review focuses on what has been done about online video peer-assessment activities. Next, this section presents studies that examined the effects of reviewers' ability on the quantity and quality of comments and review errors when the submissions were static documents. These are the variables examined in this study. Because we used several learning analytics techniques to explore reviewers' behavioral patterns, we then introduced how the learning analytics community has studied peer assessment. Finally, a theoretical framework was proposed to present the relationships between the variables examined in this study. Based on the review, we then proposed the research questions of this study.

2.1. Use of videos in online peer-assessment activities

Videos have been used as submissions in PA activities for more than 30 years. Because of the limitation of internet bandwidth, the delivery of videos in the early period was via videotape, CD, or USB. Teachers needed to make additional efforts to collect and share students' videos. With the increasing bandwidth in the recent decade, however, videos have been gradually used as submissions in online peer assessment activities. They can present dynamic visual and verbal information. Therefore, they are especially useful for evaluating actions and voices, and have been used in sport (Hsia et al., 2016b), communication skills (Lai, 2016; Lai et al., 2020), and presentations (Wu & Kao, 2008).

Studies have examined the effects of online video peer assessment activities and found that such activities can improve students' learning performance and satisfaction (Hsia et al., 2016a; Lai, 2016; Wu & Kao, 2008). For example, Hsia et al. (2016a) examined the effects of the web-based peer assessment approach on students' learning performance, self-efficacy, and satisfaction in a junior high school performing arts course. They found that, in comparison with the web-based streaming video-supported learning approach, the web-based peer assessment approach could significantly improve the students' performance and learning satisfaction. Lai (2016) implemented an online video peer assessment system for scaffolding students' communication skills. They found that students' communication performance was significantly improved. The students were satisfied with the online peer assessment learning activities.

In addition to examining the effects of the online video assessment approach, studies have developed systems and instructional approaches for supporting online video peer assessment activities (Lai et al., 2020; Lin et al., 2021; Wu & Kao, 2008). For example, Lai et al. (2020) developed a video annotation system that helped students comment on any video position. They examined the effects of the system on students' communication skills and professional attitudes during an online peer assessment activity. They found that the video annotation system was helpful for promoting students' development of communication skills, but not their professional attitudes. The students using the video-annotation tool provided more suggestion comments than those who did not use it. They concluded that the video system with the annotation function was better than the video system without the function. Lin et al. (2021) proposed an online interactive peer assessment approach with an online video peer assessment system where the interaction between assessors and assessees was two-way. They conducted an experiment to compare the approach with a one-way peer assessment approach and found that the proposed approach demonstrated significantly better learning achievement.

In sum, previous studies primarily examined the effects of online video peer assessment on performance, students' attitudes, and motivation. They found that online video peer assessment activities can improve learning performance and satisfaction. They also developed systems and instructional approaches for improving video peer assessment, and examined their effects. These developed systems and instructional approaches can provide ideas for system and instructional designers to improve online video peer assessment activities.

2.2. Peer assessment and reviewers' ability

Reviewing peers' assignments is a complex process. It consists of two intertwined tasks, providing feedback and rating. Regarding providing feedback, it involves the steps of reading, problem detection, and problem diagnosis (Patchan & Schunn, 2016). Regarding rating, it involves the steps of reading and understanding with concurrent evaluation, articulating scoring decisions, and making scoring decisions (Crisp, 2010; Cumming et al., 2002). Reviewers' ability is a key factor that may affect the reviewing process and performance.

Reviewers' ability was defined as task-related knowledge and skills (e.g., essay writing) in previous studies (Huisman et al., 2017; Xiong & Schunn, 2021). Generally, they determined reviewers' ability by a test that measured task-related knowledge and skills (Huisman et al., 2018; Patchan et al., 2013; Patchan & Schunn, 2015; Patchan & Schunn, 2016) or the quality of students' submissions (Huisman et al., 2017; Wang et al., 2016; Xiong & Schunn, 2021). For example, Patchan et al. (2013) determined the ability level of the students based on self-reported SAT verbal scores. Xiong and Schunn (2021) measured reviewers' ability by the writing quality of the submitted documents, which were evaluated by two experts. In this study, we measured reviewers' ability by the quality of their submissions, because it is most relevant to the current reviewing task (Xiong & Schunn, 2021).

When reviewing a peer's video, reviewers first watch the video. Because video is a transient medium in which the content is dynamically changing, it imposes a high cognitive load on learners (Li, 2019; Mayer, 2002). This high cognitive load may be more suitable for high-ability reviewers, because they have more prior knowledge and reserve more cognitive resources to handle the cognitive load (Moos & Azevedo, 2008; Song et al., 2016). While watching, reviewers have to concurrently detect problems. Reviewers compare their prior knowledge and the watched content to detect the problems. Because high-ability reviewers have richer knowledge of each type of problem, they should be able to easily detect problems and provide more comments (Patchan & Schunn, 2016). After detecting a problem, reviewers must provide enough information for authors to revise their submissions. A diagnosis can vary in its degree of explicitness. Providing suggestions can be seen as a more explicit diagnosis than identifying problems (Wu & Schunn, 2020). High-ability reviewers who have more knowledge of the subject and problems should be able to provide more elaborate diagnoses (Patchan & Schunn, 2015). Finally, reviewers have multiple considerations for making a final decision. The detected problems are the primary source. High-ability reviewers can effectively detect and diagnose problems. Therefore, they should be able to make more correct decisions (Xiong & Schunn, 2021).

Studies examining the effects of reviewers' ability on peer assessment activities are rare. They primarily examined how reviewers' ability affected the quantity and quality of comments that peer reviewers provide for their peers' submissions (Huisman et al., 2018; Huisman et al., 2017; Patchan et al., 2013; Patchan & Schunn, 2015; Patchan & Schunn, 2016; Xiong & Schunn, 2021). However, their results were mixed (Huisman et al., 2017; Patchan & Schunn, 2015; Patchan & Schunn, 2016). Several studies have found that reviewers' ability did not affect the quantity of comments provided. For example, Patchan and Schunn (2016) found that the number of comments of high- and low-ability reviewers was not significantly different. However, high-ability reviewers provided more high-level comments than low-ability reviewers provided. Huisman et al. (2017) also found that reviewer ability did not affect the provided feedback quantity. However, higher ability reviewers provided more suggestions and explanatory feedback than low-ability reviewers.

Patchan et al. (2013) found that high-ability reviewers provided more comments than low-ability reviewers. They examined how ability pairing (e.g., a high-ability reviewer with a high-ability author) affected the quantity and quality of comments. They found that high-ability reviewers provided more feedback, and their feedback was more likely to be implemented than that of low-ability reviewers. In particular, high-ability reviewers provided more problems, low prose issues, and substance issues for low-ability writers than low-ability reviewers. Low-ability reviewers provided more positively emotional comments than high-ability reviewers on high-ability submissions. Although the experimental results obtained in these studies differed slightly, they reported one consistent result, namely, that high-ability reviewers provided more high-level feedback than low-ability reviewers.

In addition to the comments provided, we found only one study that examined the relationship between reviewers' ability and review error. Xiong and Schunn (2021) examined the relationships between the factors related to reviewer, essay, and reviewing process and whether the factors could predict two types of review errors: severity and leniency. They defined review error as the discrepancy between peer reviews and expert reviews. Review errors were calculated using the difference between students' ratings and expert ratings on a given essay. Review errors were further categorized as severe and lenient. Their study found that reviewers' ability could predict severe errors, but could not predict lenient errors. In particular, reviewers' ability was found to be negatively related to severe errors, and lower ability reviewers were more likely to produce severe ratings. These results indicated that reviewers' ability can significantly affect review error. In our study, the definition and measurement of reviewer error is the same as the definition and measurement used in Xiong and Schunn (2021).

2.3. Peer assessment in learning analytics

Learning Analytics (LA) is a field that offers tools and techniques to analyze educational data in order to understand the process of learning and improve the education environment. Previous studies in applying learning analytics to support peer assessment activities have focused on several areas, such as learning analytics dashboards (Er et al., 2021), automatic feedback (Cavalcanti et al., 2021; Shibani et al., 2019; Shibani et al., 2022), automatically classifying reviewers' comments (Dood et al., 2022), and predicting review errors (Xiong & Schunn, 2021). For example, Er et al. (2021) proposed a theoretical framework of collaborative peer feedback and designed a learning analytics dashboard based on the framework. The dashboard, which provides an overview of participation in assessments, class-wide statistics about feedback, and an overview of several engagement indicators, aims to support instructor actions for pedagogical decisions in a peer assessment activity. Shibani et al. (2022) introduced a writing analytics tool which used natural language processing to automatically identify rhetorically salient structures in writing. The tool can then provide contextualized automated writing feedback for students' assignments. Students revised their assignments based on both automated and peer feedback.

In addition to supporting peer assessment activities, studies have applied LA techniques and tools to explore learners' behavior patterns (Er et al., 2021; Hsu et al., 2018) for peer assessment activities. Clustering analysis and sequential behavior analysis are frequently used techniques. Clustering analysis was commonly used for exploring unanticipated trends or patterns (Cerezo et al., 2016; Li & Tsai, 2017). For example, Mirriahi et al. (2016) used clustering analysis on the behavioral variables (e.g., number of annotations, video watching time, and number of pauses) of a video annotation tool used for a video peer assessment activity. They found that students' viewing behaviors showed great variety and were clustered into four behavior patterns: minimalists, task-oriented, disenchanted, and intensive users. They then found that these behavior patterns were affected by external factors (e.g., grading). Sequential behavior analysis was used for exploring the behavior transitions (Li et al., 2022; Zarzour et al., 2020). For example, Chen et al. (2020) used sequential analysis to explore students' behavioral sequences in three online video peer assessment activities: comment only, scoring only, and comment with scoring. They then compared the differences in students' behavioral patterns among the three activities. They found that the students in the comment with scoring group had better musical theater performance, provided more critical feedback, and performed more behaviors of reading the rubrics, watching example videos, watching peers' work, and reading peers' feedback.

2.4. Theoretical framework

According to the above discussions, this study aimed to examine the effects of reviewers' ability on reviewing process and performance. The Presage-Process-Product (3P) model (Biggs, 1987) was applied as a theoretical framework. This model identifies three sections: presage, process, and product. The presage section considers pre-existing individual characteristics (e.g., gender, ability, and prior knowledge) and contextual issues (e.g., learning activities, instructor effects, and learning systems). The process variables are the ways in which learners handle their learning tasks. They are the results of the interaction between individual characteristics and contextual factors. Because learners with different individual characteristics have different perceptions of their contexts, these perceptions affect their choices regarding learning behaviors and strategies. Finally, the product section includes the learning outcomes of each learner (Cybinski & Selvanathan, 2005). In this study, the presage factor is reviewers' ability; the product factors are the quantity and quality of comments provided and the reviewing error. The process factor is the behavioral pattern acquired by analyzing the system logs. The behavioral analysis may clarify the role of reviewers' ability in peer assessment activities (Chen et al., 2020; Topping, 1998). Based on this framework, this study aimed to answer the following four research questions.

- Did the reviewers’ ability affect the scores they gave?
- Did the reviewers’ ability affect their review error?
- Did the reviewers’ ability affect the quantity and quality of the comments they provided?
- Did the reviewers’ ability affect their behavioral patterns?

This study focuses on the effects of reviewers’ ability on rating error, comments provided, and behavior patterns in a video peer assessment activity. There are three reasons for this focus. First, videos have been gradually used as submissions in an online peer assessment activity. However, their transient nature imposes a high cognitive load on students, particularly for low-ability students. Reviewers’ ability is a key factor that may affect the reviewing process and performance. However, we have not found any study that has examined the effects of reviewers’ ability in video peer assessment activities. Second, previous studies that examined the effects of reviewers’ ability on the quantity and quality of comments provided revealed mixed results. In addition, we found only one study that examined the relationship between reviewers’ ability level and review error. Therefore, more research should be conducted to provide more empirical findings. Third, because reviewers can freely control their pace and path when reviewing peers’ videos in online peer assessment systems, their reviewing outcomes, such as comments provided and scores rated, are influenced by a range of factors. Understanding how reviewers’ ability affects their reviewing process and performance can help instructional and system designers to improve the system and instructional design and to design personalized supports for reviewers with different ability levels (Li & Tsai, 2020; Wang et al., 2016).

3. Method

3.1. Participants and course

This study was conducted by a quasi-experimental design. A total of 38 first-year undergraduate students (20 males and 18 females) participated in this study. They were film design majors enrolled in a one-semester course called digital editing at a university of northern Taiwan. They attended face-to-face classes, where the course teacher introduced storytelling, digital editing skills, and film editing software for 2 hours each week in a computerized classroom, in which each student used one computer with internet access. In addition to lectures in the classroom, the course teacher published peer assessment assignments on a video peer assessment system. To meet the ethical requirements, before conducting the peer assessment activity, the students were informed of the purposes of the study and read the consent letter to confirm their rights in this study. The students who had signed the consent letter were involved in the study.

3.2. Video peer assessment system

The video peer assessment system is a subsystem of a learning management system (LMS). It consists of three components: submitting, reviewing, and sharing. Teachers can create a video assignment using the submitting component. Each video assignment is presented on a submitting page where the students can upload their videos.

Teachers can use the reviewing component to create a reviewing assignment. When creating a reviewing assignment, teachers have to select a video assignment and an evaluation rubric. The rubric, which the teacher previously created using the LMS, was used by the reviewers for evaluating the assigned videos. This provides flexibility that allows teachers to assign different rubrics for different video assignments. The system then randomly assigns two peers as reviewers for each submitted video and automatically creates a reviewing page for each reviewer to review peers’ videos.

The review was anonymous. When reviewing the assigned videos, a reviewer can link to the reviewing page. Figure 1 is the reviewing page. The page presents a rubric link, which is associated with a rubric page, and a video link, which is associated with a video page, for each assigned video. On the rubric page, the evaluation rubric, which was selected by the teacher when created a reviewing assignment, is presented. A reviewer can evaluate the assigned video by the rubric. On the video page, a video annotation interface is presented (Figure 2), where the reviewer can view the assigned video and comment on any position of the video timeline.

This video annotation interface allows reviewers to add a comment at any position of the video timeline. To add a comment at a specific position, a reviewer first drags the timeline to the position. Next, he/she clicks the right mouse button and then a menu with an “Insert a comment” button is displayed. The reviewer clicks the button and then a dialog is immediately presented. The reviewer can type his/her comments into the dialog and click the submit button; then a comment tag (red rectangle) is immediately added at that position.

Figure 1. The interface of the reviewing page

Figure 2. The video annotation interface

One feature of the video annotation interface is in-context comments. The interface associated the comments and timeline, so users can easily identify how many comments have been created and where they are, and quickly view the comments and the associated video content.

Finally, teachers can share the results of a reviewing assignment by the sharing component. When a reviewing assignment is shared, a sharing page is immediately generated. The page lists all students’ videos with reviewing results, including links to the corresponding rubric pages and video pages. Authors can view the reviewing results of their videos to revise their submitted videos accordingly and learn from peers’ videos and reviewing results.

3.3. Procedure

In the first 5 weeks of the course, teachers not only introduced the concepts, skills, and software of film editing, but also used one or two examples, which were the videos submitted by the students of the previous year, to teach the students how to rate the videos, how to provide comment for the videos, and how to use the video peer assessment system in each week. A peer assessment activity was implemented during week 6 to week 8 of the course. The 1st (week 6), 2nd (week 7), and 3rd (week 8) weeks were for submitting, reviewing, and revising, respectively. At the beginning of the 1st week, the course teacher published a video assignment requiring students to edit a video and submit it to the system within 1 week.

At the beginning of the 2nd week, the teacher published a reviewing assignment where each student was assigned two peers’ videos for reviewing. The students were required to finish their reviews within 1 week. At the beginning of the 3rd week, the teacher shared the reviewing results. He also published a video assignment which required the students to submit their revised video and a document on which the students responded to the peers’ comments before the end of the 3rd week. A student was rewarded with a 16%, 8%, and 8% portion of the final grade for the quality of the submitted video, the quality of the comments provided for peers’ videos, and the quality of the revised video, respectively.

3.4. The evaluation rubric

The rubric used for evaluating the students' videos was designed by the course teacher and a film editing expert who had taught film editing for 3 years. The course teacher collected three evaluation rubrics used in video editing competitions and discussed them with the expert to determine the dimensions, detailed descriptions of the dimensions, and the rating scheme. There are three dimensions, namely rhythm, creativeness, and technical skill. The raters gave a score of 1 to 5 to every submitted video on each dimension. A higher score represents higher video quality. The detailed descriptions of the rating scheme are listed in Table 1.

Table 1. The detailed descriptions of the evaluation rubric

Dimensions	Excellent(5)	Good(4)	Average(3)	Partial(2)	Unsatisfactory(1)
Rhythm	The rhythm of the film is comprehensive, accurate, and persuasive.	The rhythm of the film is good and persuasive.	The rhythm of the film is not comprehensive and /or persuasive.	The rhythm of the film is incomplete.	The film is not presented with rhythm at all.
Creativeness	The film shows excellent ideas that can be understood by the audience.	The film shows good ideas that can be understood by the audience.	The film does not show good ideas, but the audience can understand the content.	The film does not show good ideas, and the audience can only partially understand the content.	The film does not show good ideas, and the audience cannot understand the content.
Technical skills	The film was edited with excellent quality editing skills and fully presented proper video effects and volume.	The film was edited with good quality editing skills and partially presented proper video effects and volume.	The film was edited with general quality editing skills and presented no proper video effects and volume.	The film was edited with poor quality editing skills and presented a few improper video effects and volume.	The film was edited with very poor quality editing skills and presented improper video effects and volume.

3.5. Data collection and analysis

In order to answer the four research questions, there were three kinds of data collected in this study, consisting of (1) the results of the teacher's ratings and peer reviewers' ratings, (2) peer reviewers' comments, and (3) system logs.

3.5.1. Reviewers' ability

A student's ability was determined by his/her submitted video. Each submitted video was rated by the course teacher and the film editing expert. The two raters independently rated 30% of the submitted videos based on the evaluation rubric. Cohen's Kappa analysis was performed to assess the inter-rater reliability of the two raters. The coefficients were 0.71, 0.68, and 0.72 for rhythm, creativeness, and technical skill respectively, showing that there was a high degree of consistency between the two raters. Finally, the course teacher evaluated the rest of the submitted videos. A median split was used to determine which students had higher ability and which had lower ability. Because one student did not review peers' videos, we excluded him from our analysis. Finally, the low-ability and high-ability groups comprised 19 and 18 students, respectively. The high-ability reviewers gained significantly higher scores in the rhythm dimension (U = 336.000, z = 5.206, p = .000, r = 0.856), creativeness dimension (U = 334.500, z = 5.107, p = .000, r = 0.840), and skill dimension (U = 309.000, z = 4.376, p = .000, r = 0.720) of the evaluation rubric and the sum of the three dimensions (U = 342.000, z = 5.233, p = .000, r = 0.861) than low-ability reviewers did (see Table 2).

Table 2. The scores of the high-ability group and low-ability group's video products

Indicators	Low-ability reviewers (n = 19)		High-ability reviewers (n = 18)		Mann-Whitney test
	Median	25/75 percentiles	Median	25/75 percentiles	p
Rhythm	2.00	1.00/2.00	4.00	3.00/4.00	.000
Creativeness	2.00	1.00/2.00	4.00	3.00/4.00	.000
Skill	2.00	1.00/3.00	4.00	3.00/4.00	.000
Sum of the three subscales	6.00	5.00/7.00	11.50	10.00/12.25	.000

3.5.2. Reviewing score

In addition to the teacher's ratings, each reviewer had to rate two peers' videos. The reviewing score of a peer reviewer is the average score of the two videos rated by the reviewer. There were four indicators generated from the reviewing scores. Three indicators, RhythmMeanScore, CreativenessMeanScore, and SkillMeanScore, are the reviewing scores of the three subscales respectively; and one indicator (TotalMeanScore) is the sum of the reviewing scores of the three subscales. These indicators were used for answering research question one.

3.5.3. Review error

The review error of a peer reviewer's rating for a video is the discrepancy between the scores of the course teacher and the peer reviewers. A lower review error represents higher review accuracy (Xiong & Schunn, 2021). There were four indicators generated from the review errors. Three indicators, RhythmError, CreativenessError, and SkillError, are the review errors of the three subscales respectively; one indicator (TotalError) is the sum of the review errors of the three subscales. These indicators were used for answering research question two.

3.5.4. Peer review comment coding

The reviewers' comments were qualitatively analyzed. The course teacher and the first author collaboratively developed a coding scheme (see Table 3) based on the previous studies (Cheng et al., 2015; Lu & Law, 2012). They then independently evaluated 20% of the comments based on the coding scheme. The inter-rater agreement between the two raters was calculated using Cohen's Kappa analysis, indicating a good reliability of 0.85, which is considered high agreement. Finally, the first author analyzed the rest of the comments. There were seven indicators generated from the reviewers' comments, consisting of the number of negative comments (NumNegative), number of positive comments (NumPositive), number of affective comments (NumAffective), number of comments identifying problems (NumIdentifyingProblems), number of suggestion comments (NumSuggestion), number of cognitive comments (NumCognitive), and number of all comments (NumComment). These indicators were used for answering research question three.

Table 3. Coding scheme for reviewers' comments

Categories	Definition	Example
Affective		
Negative	Giving criticism	The quality is bad.
Positive	Praising the work	Very good
Cognitive		
Identifying problems	Proposing specific problems	The video effect is not naturally presented.
Suggestions	Providing suggestions for dealing with a problem	The beginning of this video can be cut by one second to make the actors' action look smoother.

3.5.5. Behavior pattern analysis

The reviewers' operations in using the system were recorded in system logs. Generally, each recorded operation comprised four attributes: userId (who raised the operation), videoId (which video was viewed), operationName (the name of the operation, such as opening a Web page, closing a Web page, pausing, playing, adding a comment, mouse focusing on a Web page, and mouse focusing out of a Web page), and dateTime (the date and time of the operation performed). In this study, each reviewer needed to watch two videos on two video pages and rate two rubrics on two rubric pages. A reviewer performed different behavior patterns while watching videos and accessing the four pages. This study used k-means clustering analysis to explore the students' behavior patterns of watching the videos, and used lag sequential analysis to explore the behavior patterns of accessing the four pages (Hsu et al., 2018; Sun et al., 2018). These analyses intended to answer research question four.

A reviewer may open a video page several times and perform different behavior patterns on each opened page. In this study, the reviewers opened 203 video pages during the reviewing process. In order to understand the reviewers' viewing patterns on these opened video pages, this study established five variables for each opened video page and performed k-means clustering analysis on the five variables. The variables consisted of the time that the video was played (PlayTime), the time that the video was paused (PauseTime), the number of forward operations (NumForward), the number of backward operations (NumBackward), and the number of comment operations (NumCommentOperation). The five indicators were created because they are the most representative factors for actively viewing videos. It should be noted that the system cannot detect whether a student is actually on task. Students' inactivity (breaks, distractions etc.) could occupy a significant amount of time. Therefore, this study used time-oriented heuristics to place a threshold (4 min) (Kovanovic et al., 2015; Li & Tsai, 2017). The reason that we placed the threshold at 4 minutes is that the longest of the student's videos was 4 minutes. If a video was paused for a period of time longer than the threshold, the measured time was replaced with the threshold value.

K-means cluster analysis was performed on the five variables. Before doing the analysis, the five variables were transformed in order to reduce the bias in the cluster analysis (Li, 2019; Lust et al., 2011). The 0~20%, 21~40%, 41~60%, 61~80%, and 81~100% time durations or numbers were allocated a value of 1, 2, 3, 4, and 5, respectively, indicating very low, low, moderate, high, and very high. Two clusters were identified. The reviewers who spent more PlayTime and PauseTime and performed more NumForward, NumBackward, and NumCommentOperation in the opened video pages were classified into cluster 2, while those who spent less time were classified into cluster 1 (see Table 4). Therefore, Cluster 1 was labeled as "low active session" and Cluster 2 was labeled as "high active session."

Table 4. Cluster analysis of the opened video pages

	Low Active session Cluster1 (n = 107)	High Active session Cluster2 (n = 96)
PlayTime	2.084	3.990
PauseTime	2.206	3.938
NumForward	1.963	3.844
NumBackward	1.523	4.073
NumCommenOperation	1.252	3.594

To explore the reviewers' behavior patterns of accessing the four pages, we created nine codes and used lag sequential analysis to examine the patterns of accessing the four pages. The coding scheme is listed in Table 5.

Table 5. The coding scheme of the reviewers' reviewing behaviors

Code	Description
Start	Starting the peer assessment activity
End	Finishing the peer assessment activity
Break	More than one hour break between two operations.
LAW1	Performing low active session on first video page
HAW1	Performing high active session on first video page
Rubric1	Viewing first rubric page
LAW2	Performing low active session on second video page
HAW2	Performing high active session on second video page
Rubric2	Viewing second rubric page

3.5.6. Statistical analyses

This study focused on between-group (high- vs. low-ability reviewers) differences in these indicators. Therefore, group comparison methods had to be conducted. SPSS software was used for analyzing the data. Because all of the indicators violated the assumption of normality, as assessed by the Shapiro–Wilk test (p < .05), Mann-

Whitney nonparametric tests were used for the indicators. The effect size was estimated by Cohen's *r* ($r = z/\sqrt{n}$), with 0.1, 0.3, and 0.5 corresponding to small, medium, and large effect sizes (Fritz et al., 2012).

4. Results

4.1. The reviewing scores

Four Mann-Whitney tests were conducted to compare the reviewing scores of the low- and high-ability reviewers. The results revealed that the low-ability reviewers rated RhythmMeanScore (U = 81.000, z = -2.790, p = .006, r = 0.459), CreativenessMeanScore (U = 106.000, z = -2.023, p = .049, r = 0.333), SkillMeanScore (U = 88.500, z = -2.584, p = .011, r = 0.425), and TotalMeanScore (U = 80.000, z = -2.776, p = .005, r = 0.456) significantly higher than the high-ability reviewers did (see Table 6).

These results may be caused by the difference in the quality of the videos reviewed by high-ability and low-ability reviewers. To rule out the possibility, we compared the scores of the course teacher's ratings to the videos that were assigned for low-ability reviewers and high-ability reviewers. The results did not demonstrate any significant difference on the three subscales and the sum of the three subscales. These results may represent that the quality of the videos reviewed by low- and high-ability reviewers was similar.

Table 6. Reviewing scores of the low- and high-ability reviewers

Indicators	Low-ability reviewers (n = 19)		High-ability reviewers (n = 18)		Mann-Whitney test
	Median	25/75 percentiles	Median	25/75 percentiles	p
RhythmMeanScore	7.00	6.00/9.00	6.00	5.00/7.00	.006
CreativenessMeanScore	7.00	6.00/8.00	6.00	5.00/6.25	.049
SkillMeanScore	7.00	6.00/9.00	6.00	5.00/6.00	.011
TotalMeanScore	22.00	18.00/24.00	18.00	15.75/19.00	.005

4.2. The review errors

Four Mann-Whitney tests were conducted to compare the review errors of the low- and high-ability reviewers. The results revealed that the low-ability reviewers had significantly higher SkillError (U = 92.000, z = -2.471, p = .016, r = 0.406) than the high-ability reviewers did. However, the other indicators did not demonstrate any significant differences (see Table 7).

Table 7. Review errors of the low- and high-ability reviewers

Indicators	Low-ability reviewers (n = 19)		High-ability reviewers (n = 18)		Mann-Whitney test
	Median	25/75 percentiles	Median	25/75 percentiles	p
RhythmError	2.00	1.00/4.00	2.00	1.00/3.00	.518
CreativenessError	3.00	2.00/3.00	2.00	1.75/3.25	1.000
SkillError	2.00	1.00/4.00	1.00	0.75/2.25	.016
TotalError	7.00	4.00/7.00	5.50	4.00/7.25	.199

4.3. The numbers of comments provided

Seven Mann-Whitney tests were conducted to compare the numbers of different types of comments provided by the low- and high-ability reviewers. The results revealed that the high-ability reviewers provided significantly more NumIdentifyProblem (U = 235.500, z = 1.990, p = .049, r = 0.327), NumSuggestion (U = 240.500, z = 2.148, p = .032, r = 0.353), NumComment (U = 257.000, z = 2.614, p = .08, r = 0.430) and NumCognitive (U = 281.000, z = 3.345, p = .01, r = 0.550) than the low-ability reviewers did. However, low-ability reviewers provided marginally significantly more NumPositive (U = 240.500, z = 2.148, p = .032, r = 0.353) than high-ability reviewers (see Table 8).

Table 8. The number of comments provided by low- and high-ability reviewers

Variable	Low-ability reviewers (n = 19)		High-ability reviewers (n = 18)		Mann-Whitney test
	Median	25/75 percentiles	Median	25/75 percentiles	p
NumNegative	0.00	0.00/0.00	0.00	0.00/0.00	.775
NumPositive	0.00	0.00/2.00	0.00	0.00/0.00	.081
NumAffective	0.00	0.00/2.00	0.00	1.00/0.25	.169
NumIdentifyProblem	2.00	0.00/3.00	3.00	2.00/5.00	.049
NumSuggestion	2.00	0.00/2.00	4.00	1.00/6.00	.034
NumCognitive	4.00	1.00/6.00	8.00	4.75/9.25	.002
NumComment	4.00	3.00/6.00	8.50	4.75/9.25	.026

4.4. The behavior patterns

Three Mann-Whitney tests were conducted to compare the viewing patterns of the low- and high-ability reviewers. The results revealed that the low- and high-ability reviewers opened the same numbers of video pages. However, the high-ability reviewers demonstrated significantly more high active sessions (U = 225.500, z = 1.788, p = .098, r = 0.294) and fewer low active sessions (U = 112.000, z = -1.832, p = .057, r = 0.301) than the low-ability reviewers did (see Table 9).

Table 9. Quantitative reviewing behaviors of the low- and high-ability reviewers

Variable	Low-ability reviewers (n = 19)		High-ability reviewers (n = 18)		Mann-Whitney test
	Median	25/75 percentiles	Median	25/75 percentiles	p
Low Active sessions	2.00	1.00/4.00	1.00	0.00/2.00	0.057
High active sessions	2.00	0.00/3.00	2.00	2.00/4.00	0.049
Total Video sessions	4.00	3.00/6.00	4.00	2.00/6.00	0.641

Table 10. The adjusted residual table of the low-ability reviewers

	Start	LAW1	HAW1	Rubric1	Break	LAW2	HAW2	Rubric2	End
Start	0	4.132*	2.823*	-0.223	-1.144	-1.826	-1.467	-1.863	-1.467
LAW1	0	1.657	-0.481	-0.265	0.343	2.543*	-1.184	-1.530	-1.775
HAW1	0	-1.691	-1.382	5.066*	-1.111	0.380	0.275	-1.102	-1.424
Rubric1	0	-0.715	-0.762	-1.615	0.689	0.031	2.937*	0.990	-0.855
Break	0	1.802	4.135*	-0.872	-0.893	-1.424	-1.144	-0.601	-1.144
LAW2	0	-0.450	-1.772	-0.502	1.170	-1.090	-0.426	1.759	1.675
HAW2	0	-2.366	0.275	-0.223	-0.120	-1.826	0.191	4.345*	0.191
Rubric2	0	-2.022	-1.102	-0.588	0.251	1.759	0.206	-1.793	4.345*
End	0	0	0	0	0	0	0	0	0

Note. *p < .05.

Table 11. The adjusted residual table of the high-ability reviewers

	Start	LAW1	HAW1	Rubric1	Break	LAW2	HAW2	Rubric2	End
Start	0	1.712	3.683*	0.942	-1.478	-0.323	-1.826	-1.826	-1.395
LAW1	0	-0.900	1.207	0.540	2.242*	-1.375	-0.020	-1.332	-0.711
HAW1	0	-1.381	-1.932	3.593*	-1.313	0.573	3.195*	-2.438	-1.169
Rubric1	0	0.540	-0.411	-2.060	0.667	0.740	1.712	0.189	-0.932
Break	0	0.701	3.316*	-0.525	-1.567	1.352	-1.935	0.077	-1.478
LAW2	0	1.257	-0.933	-1.297	-0.442	0.899	-0.889	1.402	0.617
HAW2	0	-0.676	-1.875	-1.841	0.077	-0.889	-0.676	6.179*	-0.420
Rubric2	0	-0.676	-1.875	0.697	1.418	-0.889	-0.676	-2.390	5.204*
End	0	0	0	0	0	0	0	0	0

Note. *p < .05.

To explore the differences in the sequential patterns of the low- and high-ability reviewers, two lag sequential analyses were performed. The adjusted residual tables of the low- and high-ability reviewers are shown in Table 10 and Table 11 respectively, where the row presents the starting behavior and the column presents the following behavior. The value in each cell of the tables is the Z-score. The significant relationship is marked with a " * " when the Z-score is greater than 1.96. Figure 3 and Figure 4 further present the behavioral transition diagrams of low-ability and high-ability reviewers, respectively. They shows significant behavioral patterns. Some similar behavioral patterns were exhibited by the low- and high-ability reviewers. These are Start->HAW1, HAW1->Rubric1, Break->HAW1, HAW2->Rubric2, and Rubric2->End. In addition, several patterns differed between the low- and high-ability reviewers. The low-ability reviewers frequently performed the following patterns: Start->LAW1, LAW1->LAW2, and Rubric1->HAW2, while the high-ability reviewers frequently performed the following patterns: LAW1->Break and HAW1->HAW2.

Figure 3. Behavioral transition diagram of low-ability reviewers
Note. The arrows refer to the direction of the sequential transfer. The value on an arrow is the z-score of the sequential transfer.

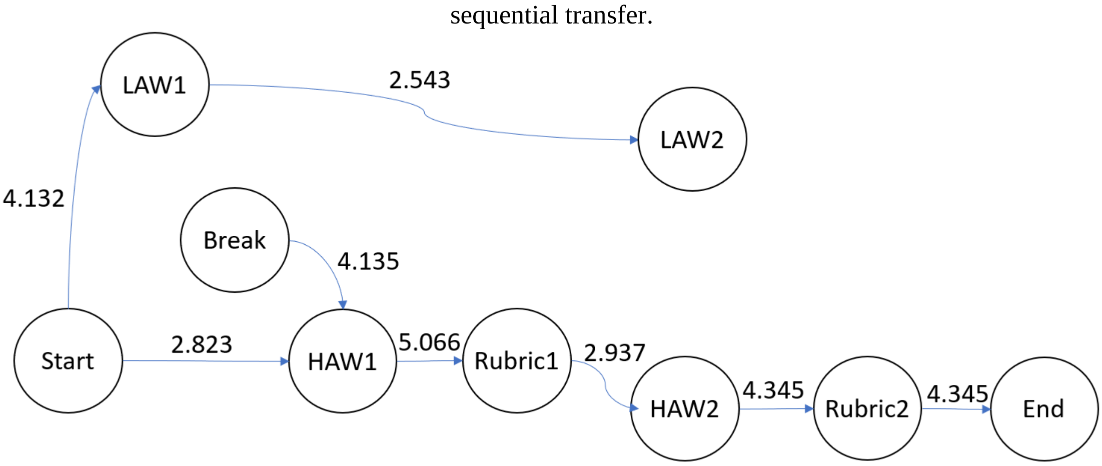
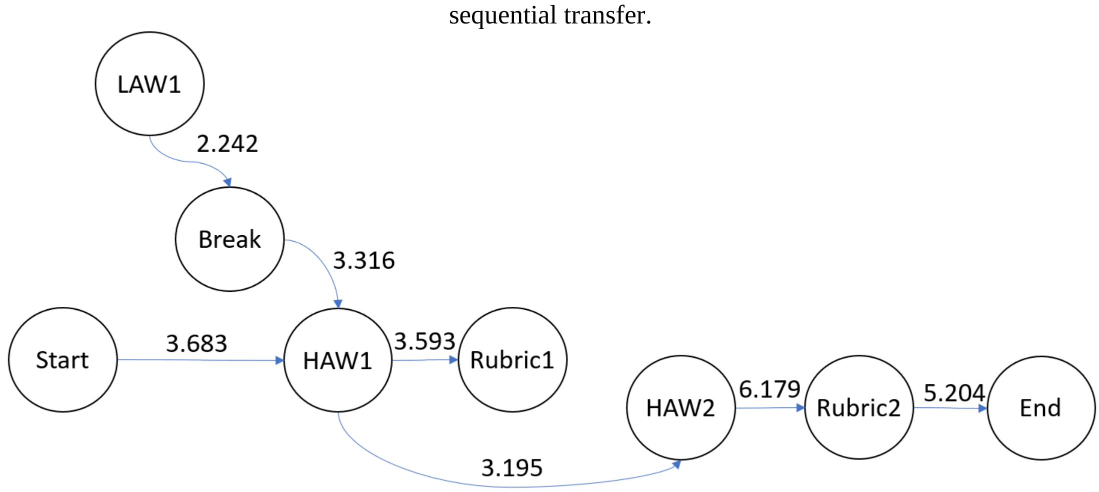


Figure 4. Behavioral transition diagram of high-ability reviewers
Note. The arrows refer to the direction of the sequential transfer. The value on an arrow is the z-score of the sequential transfer.



5. Discussion

There were four research questions in this study. For Question 1, "Did the reviewers' ability affect the scores they gave?", this study found that low-ability reviewers preferred to rate higher scores for their peers' videos than high-ability reviewers did. This result may be explained by Dunning-Kruger effect (Biango-Daniels & Sarvary, 2021; Kruger & Dunning, 1999). Lower-ability reviewers may overestimate their own knowledge or competence in film editing. When a low-ability reviewer assessed a peer's video that has the same quality as his/her video. He/She may rate a high score. The result is similar to previous studies that found peer-assessment was overestimated compared to instructors' assessment (Biango-Daniels & Sarvary, 2021; Lynch & Schmid, 2017). However, this result is inconsistent with Xiong and Schunn's (2021) study. They found that low-ability reviewers tended to be more severe. The inconsistent result may be that the dependent variables and the statistical methods were different. In our study, the dependent variable was reviewing score, a continuous data type, and was tested by Mann-Whitney nonparametric test. However, in Xiong and Schunn's (2021) study, the dependent variable was review error, which is a categorical data type. They first calculated the review errors using the difference between peer reviewers' and course teacher's ratings. They further categorized the review errors as Severe, Lenient, and Accurate. Review errors below -1 were categorized as Severe; errors above 1 were categorized as Lenient; and everything between (including) -1 and 1 was categorized into the Accurate category. They then used logistic regression to examine the relation between reviewers' ability and each review error type. Therefore, further research should be conducted to ensure the effect of reviewers' ability on reviewers' rating preferences.

For question 2, "Did the reviewers' ability affect their review error?", we found that the review errors of the subscale "technical skill" of the low- and high-ability reviewers demonstrated a significant difference, but the review errors of the subscales "rhythm" and "creativeness" did not. Because high-ability reviewers had more knowledge of the problems related to technical skills, they could more accurately identify the problems related to technical skills and had lower review errors concerning technical skills than low-ability reviewers did. However, it is not clear why the subscales of rhythm and creativeness did not demonstrate significant differences. The reason may be that rhythm and creativeness knowledge is tacit and subjective. Tacit knowledge is accumulated from immense histories of life and work experience (Tee & Karney, 2010). Students have less experience of film editing and therefore may have difficulty evaluating rhythm and creativeness knowledge.

For question 3, "Did the reviewers' ability affect the quantity and quality of the comments they provided?", high-ability reviewers provided more cognitive comments and total comments on their peers' video assignments than low-ability reviewers did. This result is consistent with previous studies (Patchan et al., 2013). However, low-ability reviewers provided more positive comments than high-ability reviewers did. Because high-ability reviewers had more knowledge of the subject and problems, they could more easily detect problems and provide more elaborate diagnoses (Patchan & Schunn, 2016). Therefore, they provided more cognitive comments. Reviewing peers' videos was a learning activity. Although the low-ability reviewers may have had difficulty identifying problems and providing suggestions, they still had to provide comments on their peers' videos. Emotional comments are more easily created than cognitive comments. Therefore, they posted more positive comments than high-ability reviewers did. These results are consistent with previous studies (Alqassab et al., 2018; Patchan & Schunn, 2015) which found that low-ability reviewers preferred to give emotional comments, while high-ability reviewers preferred to give cognitive comments.

For questions 4, "Did the reviewers' ability affect their behavioral patterns?", we found that the low- and high-ability reviewers showed different patterns of viewing peers' videos and navigating the four pages. In terms of the patterns of viewing videos, low-ability reviewers performed significantly more low active sessions than high-ability reviewers did. We have examined what reviewers did in the low active sessions and found two primary behavior patterns: long playing and short playing with a few forward and backward operations. The two viewing patterns were also observed by previous studies (de Boer et al., 2016). When the reviewers performed the pattern of long playing, they played the whole video or most of the video without any other operations or just a few other operations. This pattern may represent that they watched the video to understand the video content. On the other hand, students who performed short playing with a few forward and backward operations may have constructed an overview of the video content or wanted to find specific content. Therefore, this result that the low-ability reviewers performed significantly more low active sessions may imply that the low-ability reviewers invested more effort in understanding the video content than high-ability reviewers. However, the high-ability reviewers performed significantly more high active sessions than the low-ability reviewers did. In a highly active session, the reviewers watched the video for a long time and performed complex operations (e.g., adding and editing comments, moving forwards and backwards, and playing and pausing). This is similar to the strategic viewing behavior mentioned in de Boer et al. (2016). The reviewers performed these behaviors not only to understand the video content but also to detect and diagnose the problems. Therefore, this result that the high-ability reviewers performed significantly more high active sessions may imply that the high-ability reviewers invested more time and effort in detecting and diagnosing problems than the low-ability reviewers did. These results are similar to Li's (2019) study which found that low prior knowledge students spent most of the time viewing the videos for acquiring information, while the high prior knowledge students spent a considerable amount of time performing the viewing strategies for eliminating the discrepancies between their current knowledge state and the information presented in the videos.

In terms of navigational patterns, we found that the low- and high-ability reviewers performed some of the same sequential patterns and some different patterns for navigating the four pages. The low- and high-ability reviewers both performed the following patterns: start->HAW1->rubric1 and HAW2->rubric2->end. These patterns represent that they reviewed the first video at the beginning and the second video at the end of the whole reviewing process. These may imply that reviewers firstly reviewed the first assignment in the reviewing page and then reviewed the second one. This behavior is similar to the behavior of depth first processing of search result lists (Klöckner et al., 2004). Although low- and high-ability reviewers performed these similar behavior patterns, they also performed several different behavioral patterns. The low-ability reviewers performed the pattern Rubric1->HAW2. This pattern may represent that they reviewed the second video after reviewing the first video. Additionally, the low-ability reviewers also performed the pattern Start->LAW1->LAW2. Because the reviewers performed the low active sessions to understand the video content, the sequential patterns may represent that low-ability reviewers spent more time and effort understanding the video content at the beginning of their reviewing. We also observed that low-ability reviewers performed the pattern HAW1->Rubric1->HAW2->Rubric2. They provided detailed feedback and rated the videos one by one. This pattern may imply that the low-ability reviewers assessed the two videos separately. While high-ability reviewers performed the pattern HAW1-> HAW2->Rubric2. They provided detailed feedback for the two videos and then rated the scores. This pattern may imply that high-ability reviewers treated the two videos as a whole. They assessed the two videos in a summative way (Hsia et al., 2016b).

In sum, the low-ability reviewers provided fewer comments, demonstrated more low active sessions, and assessed the two videos separately. While the high-ability reviewers provided more comments, demonstrated more high active sessions, and assessed the two videos in a summative way. Two reasons may explain the different behaviors. First, low-ability reviewers may be less self-regulated learners. In this study, a student's ability was determined by his/her submitted video. The submitted video was an outcome of the peer assessment activity. It is closely linked to the three components of self-regulation: motivation, cognition, and metacognition (Trautwein & Koller, 2003). Second, low-ability reviewers were imposed a high cognitive load. The high cognitive load can significantly lower their self-regulated effort, the degree to which students can maintain motivation and persist with learning tasks (Hughes et al., 2018). Therefore, the low-ability reviewers may have a lower motivation to review peers' works, especially the submissions are videos. Compared with static documents (e.g., composition), the videos' navigational operations and transient nature can significantly increase students' cognitive load. On the one hand, the navigation operations in videos are more complex than in static documents (Leahy & Sweller, 2011). Learners move their eye focus to find a specific content in a static document. However, they drag the video timeline to a specific video frame and then move their eye focus to find a specific content in the video frame. The timeline does not provide any information cue for learners to locate a specific video frame. Therefore, learners may experience higher cognitive load and disorientation when navigating in a video than when navigating in a static document. On the other hand, the video is transient media, in which the content is dynamically changed. Learners must keep previously viewed content in working memory for comparing and integrating contents among different frames. It imposes a high cognitive load, especially the videos are long and complex (Leahy & Sweller, 2011; Leahy & Sweller, 2016). Because low-ability reviewers may be less self-regulated learners and may not persist with their review tasks, they provided fewer comments, performed simple operations (i.e., long playing and short playing with a few forward), and reviewed the two videos separately.

6. Conclusion

Reviewing peers' assignments is a complex process. It involves understanding the content, detecting and diagnosing the problems, and giving scores. Previous studies on written text have shown that reviewers' ability can significantly affect the comments and ratings they provide and their reviewing performance. Because video is a transient medium, watching videos imposes higher cognitive load than reading written text, and so reviewers' ability should have stronger effects on reviewing outcomes, behaviors, and performance. Therefore, this study examined how reviewers' ability affected their comments and ratings and their reviewing behaviors and performance. We found that low-ability reviewers tended to rate higher scores for peers' videos and demonstrated higher review errors than high-ability reviewers. In addition, low- and high-ability reviewers obviously performed different behavior patterns. In particular, the low-ability reviewers invested more time and effort in understanding the video content, while the high-ability reviewers invested more time and effort in detecting and diagnosing problems. Therefore, the high-ability reviewers also provided more comments for peers' videos, especially cognitive comments.

Although this study made a number of significant findings, several limitations should be mentioned. First, the sample size was small, which limits the extent of generalizability of the findings. In the future, we can involve more participants to examine the effects of reviewers' ability. Second, reviewers' ability measured by their submissions is an indirect measure. It may bias the research findings (Xiong & Schunn, 2021). In the future, reviewers' knowledge and skills relevant to detecting problems and providing feedback should be investigated. Third, the students were trained on how to provide comments and how to rate the videos in the first 5 weeks of the course and were provided with a detailed rubric. However, previous studies used different approaches to support participants' reviews. For example, the participants in Huisman et al.'s (2018) study were not trained. Patchan and Schunn (2016) provided their participants with a detailed rubric, including commonly-used general reviewing suggestions and specific guidelines. How the students were trained and supported may also influence the research findings (Liu & Li, 2014). Therefore, future works can examine the effects of different training approaches and reviewing scaffolding for the reviewing process and performance. Fourth, this study examined the effects of reviewers' ability on the reviewing process and performance. Other individual characteristics, such as previous experience of peer review and online learning, might also affect the process and performance of online video peer assessment (Sahan & Razi, 2020; Zou et al., 2018). However, we did not control the variables. They may bias the research findings. Future work can control these variables or investigate the main and interactive effect of these individual characteristics.

Despite these limitations, this study contributes to our knowledge regarding online video peer assessment. The study provides a first insight into the relationships between reviewers' ability and the reviewing process and performance for online video peer assessment practices. Because design is a progressive and repeated process, the findings of this study can provide useful information for improving our instruction and system for video peer assessment activities. Three practical implications can be derived from the findings. First, because low-ability reviewers prefer to give higher scores for peers' videos, it is suggested that teachers should consider students' ability to assign the same number of low- and high-ability reviewers for a video assignment in order to ensure fairness. Second, low-ability reviewers provided more emotional comments and fewer cognitive comments. Providing feedback has been found to lead to greater improvements from pre-test to post-test than receiving feedback (Patchan & Schunn, 2015). In particular, providing cognitive comments has stronger effects on reviewers' learning than providing emotional comments, because students can practice detection and diagnosis skills rather than just detection skills (Patchan & Schunn, 2015). Therefore, teachers and system designers should help low-ability reviewers post cognitive comments. Several suggestions may help them. First, teachers can train the students and provide clear guidelines for providing cognitive comments. Second, systems can detect the type of comment. If the comment is emotional comment, the system can ask the reviewer to elaborate on the comment. Machine learning can be used to identify the type of comment and to provide instant recommendations (Dood et al., 2022). Finally, low-ability reviewers performed more low active sessions to understand the video content. In order to help them understand the content, system developers can provide tools to help students understand the video content. For example, the video playing interface can show reviewers' viewing history, so reviewers can understand what they have done before and what video content they have watched. This may decrease the time of viewing the videos and allow more time for detecting and diagnosing problems. In addition, the video playing interface can also provide a noting function. The notes can remind reviewers what they have done and what they have thought before about the videos.

Acknowledgement

This work was supported by the Ministry of Science and Technology of Taiwan under contract numbers MOST108-2511-H-003-039-MY2 and MOST110-2511-H-003-031-MY2 and the "Institute for Research Excellence in Learning Sciences" of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

Alqassab, M., Strijbos, J. W., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*(1), 11-30. https://doi.org/10.1007/s10212-017-0342-0

Biango-Daniels, M., & Sarvary, M. (2021). A Challenge in teaching scientific communication: academic experience does not improve undergraduates' ability to assess their or their peers' writing. *Assessment & Evaluation in Higher Education, 46*(5), 809-820. https://doi.org/10.1080/02602938.2020.1812512

Biggs, J. B. (1987). *Student approaches to learning and studying research monograph.* Australian Council for Educational Research Ltd.

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A Systematic literature review. *Computers and Education: Artificial Intelligence, 2*, 100027. https://doi.org/10.1016/j.caeai.2021.100027

Cerezo, R., Sanchez-Santillan, M., Paule-Ruiz, M. P., & Nunez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A Case study in higher education. *Computers & Education, 96*, 42-54. https://doi.org/10.1016/j.compedu.2016.02.006

Chen, I. C., Hwang, G. J., Lai, C. L., & Wang, W. C. (2020). From design to reflection: Effects of peer-scoring and comments on students' behavioral patterns and learning outcomes in musical theater performance. *Computers & Education, 150,* 103856. https://doi.org/10.1016/j.compedu.2020.103856

Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *Internet and Higher Education, 25,* 78-84. https://doi.org/10.1016/j.iheduc.2015.02.001

Crisp, V. (2010). Towards a model of the judgement processes involved in examination marking. *Oxford Review of Education, 36*(1), 1-21. https://doi.org/10.1080/03054980903454181

Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: A Descriptive framework. *Modern Language Journal, 86*(1), 67-96. https://doi.org/10.1111/1540-4781.00137

Cybinski, P., & Selvanathan, S. (2005). Learning experience and learning effectiveness in undergraduate statistics: Modeling performance in traditional and flexible learning environments. *Decision Sciences Journal of Innovative Education, 3*(2), 251–271.

de Boer, J., Kommers, P. A., de Brock, B., & Tolboom, J. (2016). The Influence of prior knowledge and viewing repertoire on learning from video. *Education and information technologies, 21*(5), 1135–1151.

de Boer, J., Kommers, P. A. M., & de Brock, B. (2011). Using learning styles and viewing styles in streaming video. *Computers & Education, 56*(3), 727-735. https://doi.org/10.1016/j.compedu.2010.10.015

Dood, A., Winograd, B., Finkenstaedt-Quinn, S., Gere, A., & Shultz, G. (2022). PeerBERT: Automated characterization of peer review comments across courses. *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 492–499). https://doi.org/10.1145/3506860.3506892

Er, E., Dimitriadis, Y., & Gasevic, D. (2021). Collaborative peer feedback and learning analytics: theory-oriented design for supporting class-wide interventions. *Assessment & Evaluation in Higher Education, 46*(2), 169-190. https://doi.org/10.1080/02602938.2020.1764490

Er, E., Villa-Torrano, C., Dimitriadis, Y., Gasevic, D., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Eduardo Gómez-SánchezGómez-Sánchez, E., & Monés, A. M. (2021). Theory-based learning analytics to explore student engagement patterns in a peer review activity. *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 196–206). https://doi.org/10.1145/3448139.3448158

Formanek, M., Wenger, M. C., Buxner, S. R., Impey, C. D., & Sonam, T. (2017). Insights about large-scale online peer assessment from an analysis of an astronomy MOOC. *Computers & Education, 113,* 243-262. https://doi.org/10.1016/j.compedu.2017.05.019

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology-General, 141*(1), 2-18. https://doi.org/10.1037/a0024338

Hsia, L. H., Huang, I., & Hwang, G. J. (2016a). A Web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses. *British Journal of Educational Technology, 47*(4), 618-632. https://doi.org/10.1111/bjet.12248

Hsia, L. H., Huang, I., & Hwang, G. J. (2016b). Effects of different online peer-feedback approaches on students' performance skills, motivation and self-efficacy in a dance course. *Computers & Education, 96,* 55-71. https://doi.org/10.1016/j.compedu.2016.02.004

Hsu, T. C., Chang, S. C., & Liu, N. C. (2018). Peer assessment of webpage design: Behavioral sequential analysis based on eye tracking evidence. *Educational Technology & Society, 21*(2), 305-321.

Hughes, C., Costley, J., & Lange, C. (2018). The Effects of self-regulated learning and cognitive load on beginning to watch and completing video lectures at a cyber-university. *Interactive Technology and Smart Education, 15*(3), 220-237.

Huisman, B., Admiraal, W., Pilli, O., van de Ven, M., & Saab, N. (2018). Peer assessment in MOOCs: The Relationship between peer reviewers' ability and authors' essay performance. *British Journal of Educational Technology, 49*(1), 101-110. https://doi.org/10.1111/bjet.12520

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2017). Peer feedback on college students' writing: Exploring the relation between students' ability match, feedback quality and essay performance. *Higher Education Research & Development, 36*(7), 1433-1447. https://doi.org/10.1080/07294360.2017.1325854

Klöckner, K., Wirschum, N., & Jameson, A. (2004). Depth- and breadth-first processing of search result lists. *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1539-1539). https://doi.org/10.1145/985921.986115

Kovanovic, V., Gasevic, D., Joksimovic, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *Internet and Higher Education, 27,* 74-89. https://doi.org/10.1016/j.iheduc.2015.06.002

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lai, C. Y. (2016). Training nursing students' communication skills with online video peer assessment. *Computers & Education, 97*, 21-30. https://doi.org/10.1016/j.compedu.2016.02.017

Lai, C. Y., Chen, L. J., Yen, Y. C., & Lin, K. Y. (2020). Impact of video annotation on undergraduate nursing students' communication performance and commenting behaviour during an online peer-assessment activity. *Australasian Journal of Educational Technology, 36*(2), 71-88. https://doi.org/10.14742/ajet.4341

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*(6), 943-951. https://doi.org/10.1002/acp.1787

Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science, 44*(1), 107-123. https://doi.org/10.1007/s11251-015-9362-9

Li, L. Y. (2019). Effect of prior knowledge on attitudes, behavior, and learning performance in video lecture viewing. *International Journal of Human-Computer Interaction, 35*(4-5), 415-426. https://doi.org/10.1080/10447318.2018.1543086

Li, L. Y., & Tsai, C. C. (2017). Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers & Education, 114*, 286-297. https://doi.org/10.1016/j.compedu.2017.07.007

Li, L. Y., & Tsai, C. C. (2020). Students' patterns of accessing time in a text structure learning system: Relationship to individual characteristics and learning performance. *Etr&D-Educational Technology Research and Development, 68*(5), 2569-2594. https://doi.org/10.1007/s11423-020-09780-7

Li, S., Zheng, J., & Lajoie, S. P. (2022). Temporal structures and sequential patterns of self-regulated learning behaviors in problem solving with an intelligent tutoring system. *Educational Technology & Society, 25*(4), 1-14.

Lin, H.-C., Hwang, G.-J., Chang, S.-C., & Hsu, Y.-D. (2021). Facilitating critical thinking in decision making-based professional training: An Online interactive peer-review approach in a flipped learning context. *Computers & Education, 173*, 104266. https://doi.org/10.1016/j.compedu.2021.104266

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning, 17*(4), 420-432.

Liu, X. Y., & Li, L. (2014). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. *Assessment & Evaluation in Higher Education, 39*(3), 275-292. https://doi.org/10.1080/02602938.2013.823540

Lu, J. Y., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*(2), 257-275. https://doi.org/10.1007/s11251-011-9177-2

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. *Computers & Education, 57*(3), 2135-2144.

Lynch, D., & Schmid, B. (2017). Peer evaluation: Enhancing learning opportunities and reducing marking effort. *Proceedings of the Canadian Engineering Education Association (CEEA)*. https://doi.org/10.24908/pceea.v0i0.10368

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41*, 85-139.

Mirriahi, N., Liaqat, D., Dawson, S., & Gasevic, D. (2016). Uncovering student learning profiles with a video annotation tool: Reflective learning with and without instructional norms. *Etr&D-Educational Technology Research and Development, 64*(6), 1083-1106. https://doi.org/10.1007/s11423-016-9449-2

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The Role of prior domain knowledge. *Contemporary Educational Psychology, 33*(2), 270-298. https://doi.org/10.1016/j.cedpsych.2007.03.001

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The Effects of skill diversity on commenting and revisions. *Instructional Science, 41*(2), 381-405. https://doi.org/10.1007/s11251-012-9236-3

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591-614. https://doi.org/10.1007/s11251-015-9353-x

Patchan, M. M., & Schunn, C. D. (2016). Understanding the effects of receiving peer feedback for text revision: Relations between author and reviewer ability. *Journal of Writing Research, 8*(2), 227-265. https://doi.org/10.17239/jowr-2016.08.02.03

Sahan, O., & Razi, S. (2020). Do experience and text quality matter for raters' decision-making behaviors? *Language Testing, 37*(3), 311-332. https://doi.org/10.1177/0265532219900228

Shibani, A., Knight, S., & Shum, S. B. (2019). Contextualizable learning analytics design: A Generic model and writing analytics evaluations. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 210–219). https://doi.org/10.1145/3303772.3303785

Shibani, A., Knight, S., & Shum, S. B. (2022). Questioning learning analytics? Cultivating critical engagement as student automated feedback literacy. *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 326–335). https://doi.org/10.1145/3506860.3506912

Shirvani Boroujeni, M., & Dillenbourg, P. (2019). Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics, 6*, 16-33.

Song, H. S., Kalet, A. L., & Plass, J. L. (2016). Interplay of prior knowledge, self-regulation and motivation in complex multimedia learning environments. *Journal of Computer Assisted Learning, 32*(1), 31-50. https://doi.org/10.1111/jcal.12117

Sun, J. C. Y., Hwang, G. J., Lin, Y. Y., Yu, S. J., Pan, L. C., & Chen, A. Y. Z. (2018). A Votable concept mapping approach to promoting students' attentional behavior: An Analysis of sequential behavioral patterns and brainwave data. E*ducational Technology & Society, 21*(2), 177-191.

Tee, M. Y., & Karney, D. (2010). Sharing and cultivating tacit knowledge in an online learning environment. *International Journal of Computer-Supported Collaborative Learning, 5*(4), 385-413. https://doi.org/10.1007/s11412-010-9095-3

Tenorio, T., Bittencourt, II, Isotani, S., & Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A Systematic review of the literature. *Computers in Human Behavior, 64*, 94-107. https://doi.org/10.1016/j.chb.2016.06.020

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276. https://doi.org/10.2307/1170598

Trautwein, U., & Koller, O. (2003). The Relationship between homework and achievement - Still much of a mystery. *Educational Psychology Review, 15*(2), 115-145. https://doi.org/10.1023/a:1023460414243

Wang, J. H., Chen, S. Y., Chang, B., & Chan, T. W. (2016). From integrative to game-based integrative peer response: High ability versus low ability. *Journal of Computer Assisted Learning, 32*(2), 170-185. https://doi.org/10.1111/jcal.12125

Wu, C. C., & Kao, H. C. (2008). Streaming videos in peer assessment to support training pre-service teachers. *Educational Technology & Society, 11*(1), 45-55.

Wu, Y., & Schunn, C. D. (2020). From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology, 60*, Article 101826. https://doi.org/10.1016/j.cedpsych.2019.101826

Xiong, Y., & Schunn, C. D. (2021). Reviewer, essay, and reviewing-process characteristics that predict errors in web-based peer review. *Computers & Education, 166*, 104146. https://doi.org/10.1016/j.compedu.2021.104146

Zarzour, H., Bendjaballah, S., & Harirche, H. (2020). Exploring the behavioral patterns of students learning with a Facebook-based e-book approach. *Computers & Education, 156*, 103957. https://doi.org/10.1016/j.compedu.2020.103957

Zou, Y., Schunn, C. D., Wang, Y. Q., & Zhang, F. H. (2018). Student attitudes that predict participation in peer assessment. *Assessment & Evaluation in Higher Education, 43*(5), 800-811. https://doi.org/10.1080/02602938.2017.1409872