Peer Assessment of Webpage Design: Behavioral Sequential Analysis Based on Eye Tracking Evidence

Ting-Chia Hsu*, Shao-Chen Chang* and Nan-Cen Liu
Department of Technology Application and Human Resource Development, National Taiwan Normal University // ckhsu@ntnu.edu.tw // ntnuresearcher@gmail.com // vmilkv1412@gmail.com
*Corresponding authors

ABSTRACT

This study employed an eye-tracking machine to record the process of peer assessment. Each web page was divided into several regions of interest (ROIs) based on the frame design and content. A total of 49 undergraduate students with a visual learning style participated in the experiment. This study investigated the peer assessment attitudes of the participants and found that they possessed highly positive attitudes towards and understanding of peer assessment. After comparing the results of the peer assessments and evaluation by experts, high consistency occurred when the design of the web page was concise; however, the consistency decreased when the web page content was too diverse. After comparing the peer assessment attitudes of the participants and their web page design scores, it was found that the web pages with concise designs attracted the visual-style students' attention more, and that there was a significant negative correlation for those students who possessed a more negative attitude toward peer assessment. In addition, the study further analyzed the visual-style students' serial behaviors in the peer assessment process for each web page design. After comparing the evidence of each student's eye movements and his/her evaluation results, it was found that the students who gave higher or lower scores had different eye movements. For the website scored as having the best design, the fixations and behaviors of the assessors giving higher scores were highly consistent with those of the experts, implying that the few assessors giving lower scores were relatively poor at peer assessment. On the contrary, for the website which was scored as having the worst design, the fixations and behaviors of the assessors giving lower scores were highly consistent with those of the experts. Consequently, from the eye fixation hotspot evidence, when the students were more concentrated on the peer assessment, their evaluated results were closer to those of the two experts. Finally, the study found that the eye fixation hotspots were the same as the key points planned by the student designers of the website which scored the highest, which provided the student designers with additional important eye-tracking feedback from the peer assessment activities.

Keywords

Peer assessment, Eye tracking, Region of interest, Hotspots, Serial behavior

Introduction

The adoption of peer assessment has become a trend in the classroom (Boud, Cohen, & Sampson, 1999), and there is already much research indicating that peer assessment promotes the effective development of cognition and affect. However, the quality and fairness of the assessments by those students who lack experience and confidence when assessing their peers' work have been questioned by many scholars (Ballantyne, Hughes, & Mylonas, 2002; White, 2009; Seng & Hill, 2014). Previous studies have also revealed that peer assessment is influenced by differences in the students' culture, personality, and friendship (Johnston & Miles, 2004; Panadero, Romero, & Strijbos, 2013; McLeay & Wesson, 2014). Therefore, how to overcome the problem of participants' individual differences in the peer assessment process has become an important issue.

Meanwhile, a growing number of studies have been exploring the impact of individual differences on peer review. Liu, Lin, and Yuan (2002) investigated the differences in the executive thinking styles of students in the peer assessment process. Learning style is one of the individual differences in digital learning studies. Solomon and Felder's Learning style model is one of the famous models, and includes four facets (Felder & Silverman, 1988; Soloman & Felder, 2005), namely action/reflective, sensing/intuitive, visual/verbal, and sequential/global. The past learning style survey results or eye movement study results have shown that most students tend to adopt a visual learning style (Hsu, Hwang, & Chang, 2014; Hsu, Hwang, Chang, & Chang, 2013). However, one past study indicated that students with mixed learning styles constitute a higher proportion than single-style learners (Lujan & DiCarlo, 2006).

There is another important factor affecting peer assessment outcomes, namely students' attitudes. There is a positive correlation between a positive attitude towards peer assessment and higher learning achievement. Therefore, in order to ensure the validity of peer assessment, teachers need to consider the students' personal affect during the learning process (Lin, Liu, & Yuan, 2001). One past study also found that students usually have a positive attitude towards peer assessment activities. These positive attitudes ensure that the students are fair and responsible when completing peer assessment (Cheng & Warren, 1997). However, some students also worry

about their ability and their responsibility when evaluating their peers' work. To solve this problem, teachers must provide appropriate support to alleviate student stress (Cassidy, 2006).

In the last decade, many researchers have begun to analyze students' learning behavior. Because of the different learning behaviors which affect students' learning performance during the learning process (Hwang & Chen, 2016), learning behavior analytics has become an important issue in education (Hwang, Chu, & Yin, 2017). Learning behavior refers to the record of data related to students' behaviors and interactions with peers during learning activities (Hwang, Hsu, Lai, & Hsueh, 2017). Hwang, Hung, Chen, and Liu (2014) emphasized that instructors could effectively improve their teaching methods and learning activities by analyzing learning logs or educational data. However, there are few or even no studies related to analyzing behaviors during peer assessment.

Therefore, this study used quantitative and qualitative evidence to explore the results of peer assessment by students with the same kind of learning style (i.e., visual learning style). When exploring the individual behavior of students with the same kind of learning style in the website reading process, the scientific evidence of the eye tracker can be more objective than just knowing the results of the peer assessment score. Moreover, the quantitative results of peer assessment can be used to compare the differences between the learning behaviors of students who give high and low evaluation scores. This study also investigated the correlation between the goal of the students' attention and the goal (region) of the peer assessment process, and whether it was related to the students' attitudes towards peer assessment. We could then count the eye-catching hotspots, which can provide another kind of important eye movement information feedback to the web page designers.

The research questions are as follows:

- Was the assessment of the student reviewers similar to the expert appraisal? What were the student reviewers' attitudes towards peer assessment?
- What were the differences in the behavioral sequential analysis of eye movements during the peer assessment for the four websites with different designs?
- How did the behavioral patterns of the students giving peer assessment with low scores differ from those of the students giving peer assessment with high scores?

Literature review

Eyes movement tracking

The rapid development of eye tracking technology is based on the eye-mind hypothesis. Several researchers have found that one's eye gaze is closely related to attention when dealing with visual messages (Just & Carpenter, 1976). Several scholars have proposed the premotor theory to explain the relationship between eye movement and attention. The main concept of the premotor theory is concentrated on the target, and uses eye movement to aim at the target to capture the visual behaviors. That is, attention is the pre-treatment before eye movement (Hoffman & Subramaniam, 1995).

At the same time, some scholars have referred to the Noticing Hypothesis to explain the relationship between attention and cognition. Schmidt (1990) considered that learners must be aware of the message input to ensure that learning is happening. It also means that learners must concentrate on learning to acquire knowledge when they are learning (Schmidt, 1990; Schmidt, 1993; Schmidt, 1995; Schmidt, 2001). In addition, attention can determine what information will be transferred from short-term into long-term memory during the cognitive process. That is, it is clear that attention is the key to learning (Egi, Fujii, & Tatsumi, 2002). Thus, attention and eye movement are closely related, and the development of eye tracking technology provides an additional reference for exploring previously unknown areas (Godfroid, Housen, & Boers, 2010).

The process of eye movement consists of two elements, the saccade and fixation. The saccade is the action of the eye moving, and fixation is the action of the eye stopping its movement and beginning to get the message. Fixation falls within the range of 100 to 500 milliseconds, depending on the target material. Eye movement is therefore constituted of a series of gazes and jumps (Rayner, 1998; Rayner, 2009). For eye tracking studies, regions of interest (ROIs) are often used by researchers to record visual behavior. ROIs will have definitions that are appropriate for different research questions. There are four common types of eye movement data, namely duration of the first fixation (DFF), latency of the first fixation (LFF), number of fixations (NOF), and total contact time (TCT). It is also possible to further measure the course of saccades (COS) in order to investigate

students' eye movement behavior. This is the most important eye movement data collected in this study (Hewig, Trippe, Hecht, Straube, & Miltner, 2008).

In past studies, some scholars adopted eye tracking technology to develop the adaptive digital learning (Adaptive E-Learning through Eye Tracking, AdELE) framework, using it to synchronize the recording of learning content and user information. Barrios et al. (2004) investigated users' performance, knowledge level, and eye movement behavior to achieve an adaptive learning environment. In order to help the learners more, the eye tracking will record the user's eye movement behavior instantly to provide the corresponding knowledge in different areas (Calvi, Porta, & Sacchi, 2008; Gütl et al., 2005; Pivec, Trummer, & Pripfl, 2006).

Tsai, Huang, Hou, Hsu, and Chiou (2016) utilized eye-tracking technology to explore the differences between high and low conceptual comprehension of university students' visual behaviors and game flow in game-based learning. The results indicated that the students in the high-comprehension group showed efficient reading strategies for text and better metacognitive controls of visual attention during game play. Furthermore, Lindner, Eitel, Strobel, and Köller (2017) designed a science learning activity to record the eye movements of 62 schoolchildren solving multiple-choice (MC) questions. The results showed that the time the students spent fixating on the picture was compensated for by less time spent reading the corresponding text. In text-picture items, students also spent less time fixating on incorrect answer options.

Yang, Huang, and Tsai (2016) attempted to explore the effects of epistemic beliefs and the gender differences in epistemic beliefs in science on science-text reading. They administered the scientific epistemological beliefs (SEBs) questionnaire to the students and recorded the science-text reading process of the eye-tracking. The results demonstrated interactions between SEBs and gender, and the complicated SEBs were associated with higher cognitive attention to the reading of data-related information. Scholars have reviewed the eye-tracking studies and concluded seven learning-related categories employing the eye-tracking machine, namely patterns of information processing, effects of instructional strategies, re-examination of existing theories, individual differences, effects of learning strategies, patterns of decision making, and social effects (Lai et al., 2013).

From the above literature, we can see that eye tracking technology has been widely used in the field of e-learning. In addition to giving the students feedback with simple criteria items, peer assessment should pay attention to the differences between individuals. Hence, in this study, we used eye tracking technology to explore the differences between the learning styles and learning behavior of students.

Peer assessment

Peer assessment (PA) is a strategy for individuals to assess the work or learning outcomes of peers at the same level, helping students improve their cognitive ability (Topping, 1998). In the peer assessment process, all of the students not only act as the assessor, but also as the assessee (Van Zundert, Sluijsmans, & Van Merriënboer, 2010; Li, Liu, & Steckelberg, 2010). The characteristic of peer assessment is that it gives the students specific tasks and focuses on peer performance rather than on individual abilities. In this process, peers can be in the same or different classes, or even have different abilities and be in different teams when executing the peer assessment activity (Lui & Andrade, 2014). Moreover, students' understanding of their peers' ideas during the learning process is also enhanced, and their weaknesses improved (Cho & Cho, 2011).

In the past decade, much research has reported the benefits of applying peer assessment to different skills training in school settings, such as communication skills (Lai, 2016; Wang, Liang, Liu, & Liu, 2016), writing skills (Cheng, Liang, & Tsai, 2015; Ashton & Davies, 2015; Tenório, Bittencourt, Isotani, Pedro, & Ospina, 2016; Russell, Van Horne, Ward, Bettis, & Gikonyo, 2017), and higher-order thinking ability (Gielen & De Wever, 2015; Wang, Hou, & Wu, 2017). Peer assessment has been confirmed as being able to improve students' learning performance, especially their writing skills (Cheng, Liang, & Tsai, 2015). In addition to improving their knowledge, students also respect the opinions of their peers and share their experience with others in the peer assessment activity (Tayem et al., 2015; Yu & Wu, 2013). Peer assessment will not only improve the learning outcome in the course, but will also increase social interaction and extracurricular discussion outside the class. Moreover, students' high learning motivation will enhance their self-learning behavior (Xie, 2013). Peer assessment activities can also reduce the burden of teachers, provide opportunities for self-learning, and increase learning motivation and high-level thinking ability (Luo, Robinson, & Park, 2014).

Lai and Hwang (2015) proposed an interactive peer-assessment criteria development approach to help students make reflections while viewing their peers' work. A total of 103 elementary students participated in their activity

and were assigned to an experimental group and a control group. From the study results, it was found that the proposed approach significantly improved the students' learning achievement and their learning motivation. It was also indicated that engaging students in assessment criteria development is an effective approach. Moreover, Wang, Hou, and Wu (2017) designed a 4-week collaborative learning activity involving four widely used instructional strategies, namely problem solving, peer assessment, role playing, and peer tutoring. The results showed that the students exhibited relatively more instances of the cognitive process of understanding when using the peer assessment and peer tutoring instructional strategies with Blogs. In other words, peer assessment can improve students' ability to listen to suggestions and to discuss with peers.

Recently, Lai (2016) proposed an online video peer review system to support nursing students' training in communication skills. A total of 50 Taiwanese nursing students attending a Psychiatric Care program participated in the study. In the learning activity, the students were asked to treat a simulated patient and record a YouTube video to have peers watch, rate, and give feedback on it. In the study, it was found that online peer assessment can be effective in nursing communication skills training, as their communication became more patient-centered.

Although many studies have reported the effectiveness of peer assessment in the last decade, several researchers have indicated that instructors should provide specific guidelines for how to decide a rating with subjective consciousness (Tenório, Bittencourt, Isotani, & Silva, 2016; Ng, 2016; Russell, Van Horne, Ward, Bettis, & Gikonyo, 2017). Therefore, the assessment criteria in the current study leave some flexibility for the students, such as allowing them to select "maybe" for certain items. Russell et al. (2017) investigated students' evaluating processes and their perceptions of peer assessment when they were engaged in peer assessment using Calibrated Peer Review. They suggested that instructors may provide specific learning scaffolding for students of mid- or low-level ability.

Learning behavior analysis

According to Huang et al.'s (2014) definition of learning behavior analysis, students' behaviors in learning activities can be recorded to analyze the factors affecting their learning performance. In recent years, much research has reported the benefits of learning behavior analysis for improving learning activities, such as in language courses (Chen & Yeh, 2017; Hwang, Hsu, Lai, & Hsueh, 2017), social networks (Darban & Polites, 2016; Buckley & Doyle, 2017), computer science (Su, Yang, Hwang, Huang, & Tern, 2014), and science (Chiang, Yang, & Hwang, 2014).

For example, Hwang, Hsu, Lai, and Hsueh (2017) proposed a problem-based English listening game based on progressive sequential analysis to investigate the learning behavioral patterns of students with different levels of English anxiety. The experiment results indicated that high levels of English anxiety influenced students' learning achievement and gaming behaviors. Moreover, Yin, Okubo, Shimada, Oi, Hirokawa, and Ogata (2015) proposed an e-book system which recorded the students' learning logs of their daily academic life in the cloud. They found that the students' learning behaviors were significantly related to their learning performance.

In addition, Chen and Yeh (2017) developed an online test to help students learn Academic English, and investigated the effects of the students' cognitive styles in the English learning context. The results of the quantitative measurement indicated that cognitive styles have significant influences on students' learning patterns in the context of Academic English. However, there is a lack of studies investigating how learning styles affect students' reactions in the context of peer assessment. To address this issue, we designed an eye tracking activity to investigate the sequential behaviors of students with different learning styles in a peer assessment activity.

Learning styles

Learning styles are defined as the individual learning habits, methods, and attitudes of a learner, and are influenced by the learner's interactive skills and the learning environment (Dunn & Dunn, 1987). The study of learning style is derived from the cognitive style of psychology; the most important aspect of cognitive style is learning style. Therefore, the measurement of learning style should focus on the various learning activities that determine the individual's cognitive style (Sternberg & Grigorenko, 1997). Because of the rapid development of learning style research in recent years, the term learning style has become widely used.

Many scholars have begun to advocate the learning style model, because it can improve the plight of higher education teaching by helping teachers engage in more profound thinking and face the diversity of students (Claxton & Murrell, 1987). There are other scholars who also believe that in higher education it should not be assumed that all students can use the same way of learning. Teachers should be responsible for increasing their adaptive teaching methods to allow different kinds of students to achieve more effective learning (Hawk & Shah, 2007). That is, when teachers are designing courses and preparing teaching materials, they need to consider how to help students develop their interpersonal relationships, leadership, and communication skills.

Felder and Silverman (1988) proposed a learning style theory which is compatible with different learning styles and teaching methods. Based on that theory, Soloman and Felder (2005) further proposed the Index of Learning Styles (ILS), which includes the four dimensions of active/reflective, sensing/intuitive, visual/verbal, and sequential/global learning. In addition, some scholars have attempted to study the psychological model (such as human behavior and learning style) in information technology. The results showed that the Felder-Silverman learning style model has the advantages of simplicity and comprehensiveness, and is well suited for use in digital learning (Fatahi, Moradi, & Kashani-Vahid, 2016). For example, some scholars have proposed a model that examines the mediating processes in the relationship between learning style and e-learning performance. The results showed that the sensory/intuitive dimension learning style will indirectly affect learning performance through the regulation of online participation, while other types of learning styles do not have any effect (Huang, Lin, & Huang, 2012).

Therefore, this study attempted to combine learning style and eye tracking to investigate students' sequential behaviors in a peer assessment activity. Through the learning style and eye tracking sequence analysis of the records, it was hoped to understand the students' learning situation, in order to provide relevant research results and recommendations for future researchers.

Method

Participants

The study participants were undergraduates from a university in the north of Taiwan, and were approximately 20 to 21 years old. There were 80 students who had taken the e-learning course before this activity. After deducting the missing values, 57 participants with complete data remained. According to the learning style scale (Index of Learning Styles) analysis, 49 of the participants were visual learning style students (21 female and 28 male). None of these 49 students had used an eye-tracking machine before.

Measuring tools

The questionnaire of learning styles was adopted from the measure developed by Soloman and Felder (2005). It consisted of 44 items for evaluating learners' dimensions of learning style including active/reflective, sensing/intuitive, visual/verbal, and sequential/global. Each question has two opposite answers, which represent different individual learning styles. In this study, we only used the 11 visual/verbal dimension questions to measure whether or not the students adopted a visual learning style (Soloman & Felder, 2005).

Table 1. The Kelly repertory grid designed by the website experts

Different website features	Website A	Website B	Website C	Website D
The framework of the upper and lower pages.	5	1	2	1
The framework of the left and right pages.	1	5	1	5
There is no frame page.	1	1	4	1
The website interaction (Such as bulletin boards, contact boxes, information management, business-related websites and resource links, etc.)	5	1	5	1
The color coordination of the web pages.	5	5	3	1
The layout is simple.	5	1	3	5
The text is interesting.	4	1	5	5

Note. 5 = Strongly agree, 1 = Strongly disagree.

In this study, four of 10 groups of students' websites designed in the web design course were selected, each representing one of four different themes. The first website was designed to promote pearl beads and sell brand-

related products. The second website was an introduction to the school dance club organization and activities. Another website was an introduction to the famous night markets and restaurants near the campus. The last one is a website designed to sell second-hand clothing. Each website consisted of 10 web pages. The four websites were selected for the peer assessment activity, and the students scored them in accordance with the rating scale designed by the expert, as shown in Table 1.

Table 2. The rating scale of web peer assessment

Score item	Please select your rating of the students' website			
Rating title	YES (Positive)		NO (Negative)	
	Very good	Good	Bad	Very bad
The font size of the text.				
The placement of images.				
The hyperlink location design and arrangement.				
The layout will change when the user changes the page.				
The logo or trademark is obvious.				
The background of the website is appropriate.				
It is easy to find the important information on the web page.				
Do you feel that you were a concentrated reviewer during the peer assessment?				

The peer assessment scale was developed by two experienced professors who have designed e-learning digital content for several years, as shown in Table 2. The Cronbach's α value of internal consistency reliability was .846, showing that the peer assessment was highly reliable. After we finished the design of the peer assessment scale, we also scored the four web pages. The Kappa consistency score of reliability was 0.903, showing that there was a high degree of consistency between the two experts.

In this study, we modified the college students' peer assessment scale developed by Wen and Tsai (2006). The purpose was to investigate students' attitudes and acceptance of the peer assessment activities. The Cronbach's alpha value of the analysis result was .75, showing good reliability in internal consistency.

Research design

Before the test, the participants were divided into two groups according to their different learning styles and the assessment sequence to complete the peer assessment of the four websites. In order to avoid the hangover effect of peer assessment and tired eyes, we used a counterbalanced design, as shown in Table 3.

Table 3. Counterbalanced design

Random	1st time of eye tracking			2nd time of eye tracking		
Group 1	Website A	→	Website B	Website C	→	Website D
	Website B	→	Website A	Website D	→	Website C
Group 2	Website C	→	Website D	Website A	→	Website B
	Website D	→	Website C	Website B	→	Website A

Figure 1 shows the experimental design of this study. Before the learning activity, the two groups of students were instructed in digital multimedia design to learn its basic knowledge, which is a part of the learning course. Before the experiment, we asked the students to complete the learning style questionnaire. We then explained the eye tracking instructions and the eye tracking correction. During the learning process, the eye tracking machine recorded each user's process of peer assessment as they completed the peer assessment scale for the web page. After that, the students rated another website, and each website was rated two times until the students completed all of the ratings. After the learning activity, the students completed the questionnaire to measure their learning attitudes towards peer assessment.

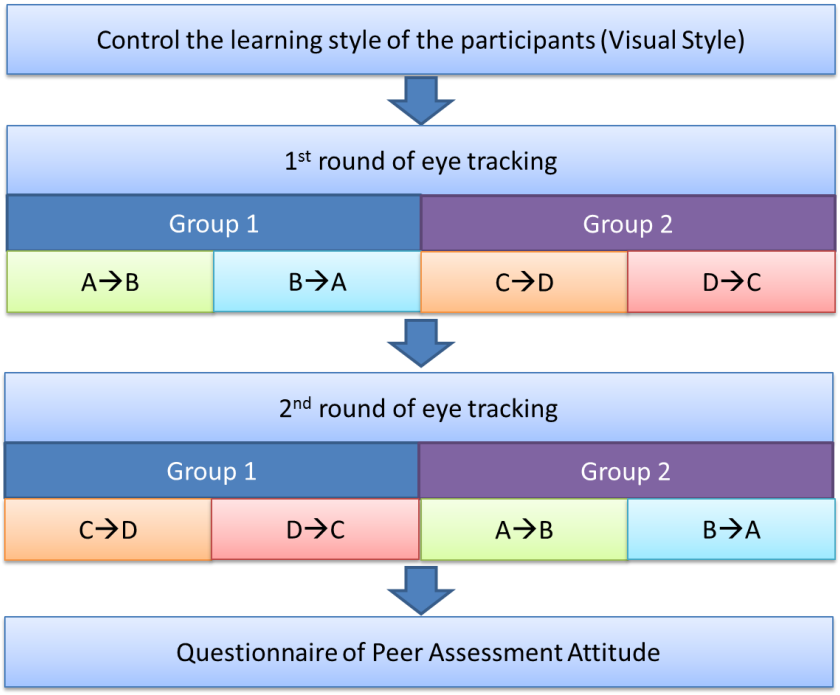


Figure 1. The process of the experimental activity

Research instruments

The research equipment used was the Eye NTNU-180 (a cheap multi-screen eye tracker) developed by the National Taiwan Normal University, Department of Electrical Engineering research team. This eye tracking instrument uses infrared LED detection of eye gaze focus on objects, and has a sampling frequency of between 30 and 180 Hz within 60 cm. The error rate is less than 0.3 degrees, and there is a 9-point correction technique to improve accuracy. An example is shown in Figure 2.

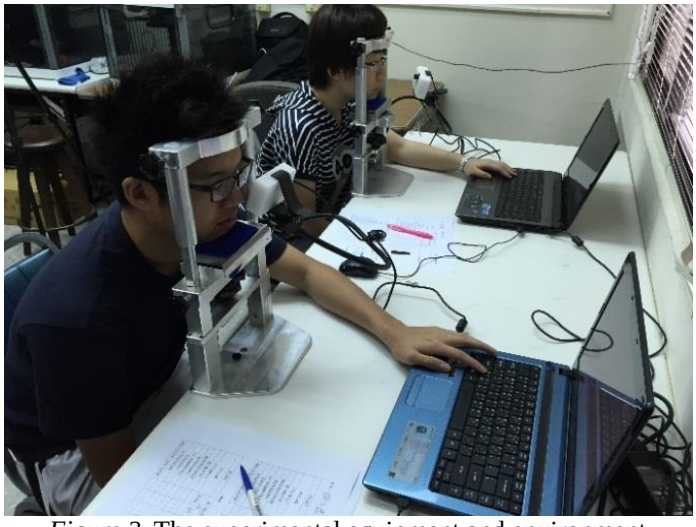


Figure 2. The experimental equipment and environment

As each web page has a frame, we categorized the text and the image area of the page frame as an ROI. We analyzed the data of the visual students in the peer assessment process and summed the scores excluding the personal attentiveness score. According to Kelley's (1939) definition, the top 27% of the students were classified as the high evaluation group, and the bottom 27% were classified as the low evaluation group (Kelley, 1939). The number of students reviewing the four websites is shown in Table 4.

Table 4. The distribution of the student reviewers

	A-Website	B-Website	C-Website	D-Website
The number of students giving the website high scores	10	13	14	14
The number of students giving the website low scores	14	11	12	13

Results

Quantitative results of peer assessment of the webpage designs

The students showed that they had concentrated on the process of peer assessment for each website. Based on the 4-point Likert scale ranging from 1 to 4, the self-report inventory showed that the reviewers' average concentration degrees for websites A, B, C, and D were 3.20, 3.04, 3.14, and 3.22 respectively during the peer assessment. In other words, the students presented high confidence in their concentration on the process of peer assessment.

The performance of the four websites was also ranked by the experts. The ranking from better to worse was A, D, B, C. After comparing the assessments of the experts and the student-reviewers, it was found that the consistency between their scoring decreased when the webpage design was worse. The similarity degree between the evaluation results of the student reviewers and the expert reviewers was more than 85% for websites A and D. Accordingly, the students' evaluation was closer to that of the experts when the webpage had a better design.

On the contrary, the evaluation of the student reviewers had about 60% similarity with the assessment of the expert reviewers for websites B and C, as shown in Table 5. Therefore, the grades given by the students were higher than those given by the professors when the webpage was not well designed. The experts gave stringent evaluations and made clear judgements, while the student reviewers seemed to be concerned about their peers' feelings. The students tended to give their peers higher grades when their proficiency was not professional enough to judge the poor design of the website, or if there was too much diversity in the design presented in one website. For example, the background of website B was so bright that the information displayed was relatively neglected.

Table 5. The consistency percentage of the experts' and student reviewers' assessments per evaluated category.

Material Website	Text	Picture	Link	Change page	LOGO	Background	Information	Mean
A	75.51%	97.96%	97.96%	87.76%	87.76%	93.88%	89.80%	90.09%
B	59.18%	51.02%	71.43%	57.14%	36.73%	87.76%	28.57%	55.98%
C	73.47%	51.02%	36.73%	44.90%	46.94%	53.06%	51.02%	51.02%
D	79.59%	97.96%	93.88%	71.43%	91.84%	75.51%	85.71%	85.13%

It was found that 27 students achieved 70%-79% similarity with the experts, which constituted the majority of the students. Secondly, 14 students had 60% - 69% similarity in comparison with the assessment of the experts. However, there was no correlation between the peer assessment attitudes of the students and the similarity degree of the students' peer assessment and the experts' evaluation, as shown in Table 6.

Table 6. The distribution of different consistency levels

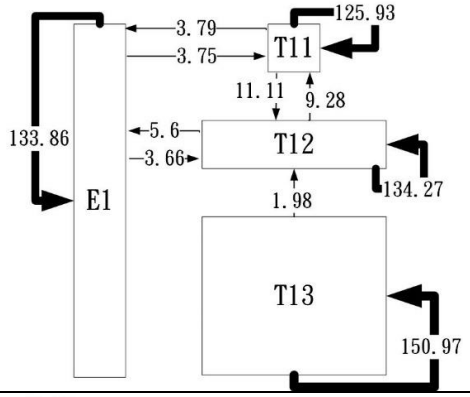
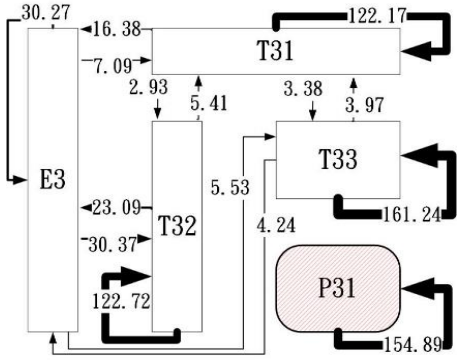
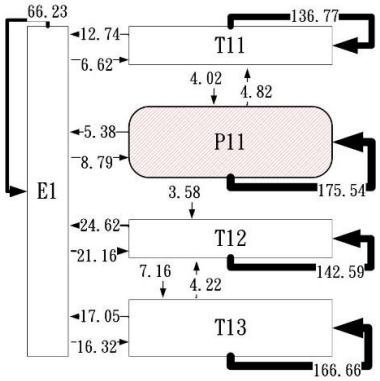
Proportion attitude	50%-59%	60%-69%	70%-79%	80%-89%	90%-100%	Mean	*SD*
N	4	14	27	3	1	49	49
Positive attitude	3.72	3.53	3.16	3.67	2.25	3.32	0.56
Understanding-and-action	3.92	3.98	3.94	3.89	4.00	3.95	0.45
Negative attitude	2.31	2.84	2.42	3.17	2.00	2.57	0.71

In terms of the peer assessment attitudes, this study utilized a one-sample *t* test to examine whether each dimension of peer assessment was significantly larger or smaller than the middle of the 5-point Likert scale (i.e., 3). The results showed that 49 students had significantly positive attitudes towards peer assessment with mean = 3.32 and *SD* = 0.56 ($t = 3.988^{***}$, $p < .001$), and showed the attitude of understanding-and-action for peer assessment with mean = 3.95 and *SD* = 0.45 ($t = 14.446^{***}$, $p < .001$). There was negative significance for negative attitudes towards peer assessment with mean = 2.57 and *SD* = 0.71 ($t = -4.252^{***}$, $p < .001$). Accordingly, the students presented high understanding and positive attitudes and showed lower negative attitudes towards the peer assessment activities.

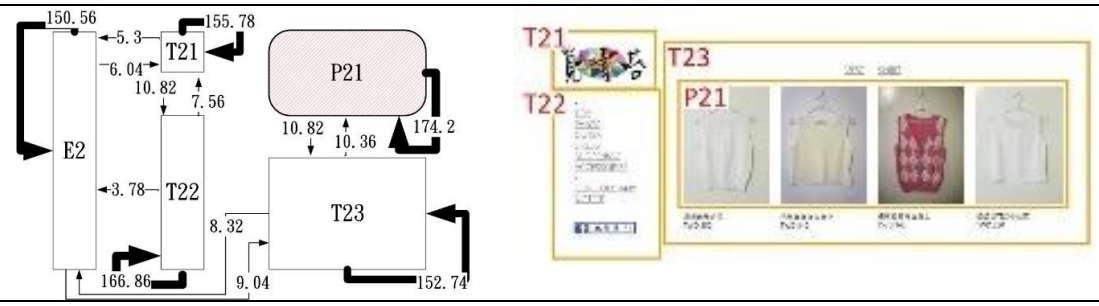
Behavioral sequential analysis of eye movements

In this study, we evaluated the students' saccade path (i.e., course of saccade, COS) among the regions of interest (ROIs) such as the text paragraph, the picture area, the menu area, the logo, or the region outside of the webpage information area. A total of 129,020 eye-tracking records were collected, with 27,224 for website A, 28,773 for website B, 36,015 for website C, and 37,008 for website D. Each webpage was divided into a maximum of five regions. After using the time serial sequential analysis method to find the significant paths ($z > 1.96$, $p < .05$) of eye movement from one region to another, each page of every website was evaluated. The arrow is thicker when the z scores between the two regions are larger. The z scores are larger when the students saccade from one region to another more frequently. For example, the behavioral sequential patterns for the first page of website A are shown in Table 7. Overall, from the results of the behavioral analysis of website A, the students put most of their focus on the three main areas. They read the logo first, then read the menu, and finally read the following information. Consistent behaviors were found for seven of the 10 webpages of website A.

Table 7. Behavioral sequential analysis of eye movements

Website	Behavioral serial patterns	Webpage template
A		
B		
C		

D



Note. Codes for the different areas: E refers to the area outside of the ROIs; P refers to a picture ROI; T refers to a text area ROI; the first number refers to the page number; the second number refers to the number of ROIs.

As for website B, there were too many COS (course of saccade) behaviors. The students spent more effort moving their eyes among the different frames. From the overall saccade behaviors of website B, too many frame pages combined in one web page did not make for a good design because the students were not able to ignore unnecessary information (e.g., the fancy background) and focus on the particular point that the designer wanted them to pay attention to.

In terms of website C, the students mainly moved their eyes between the title and the explanations. Therefore, when the two pieces of information were correlated, the students got used to matching the corresponding areas, resulting in the eye behaviors moving back and forth. One behavior was the same for websites B and C. The students scanned around the space in the web page. They may have been searching for more information.

Overall, in website D, a logo and menu which were put in a fixed place inferred that they had a more conspicuous design which was able to attract attention. The explanations had best not be too far from the pictures so that the students could easily scan between the pictures (e.g., the clothes pictures) and their corresponding explanations.

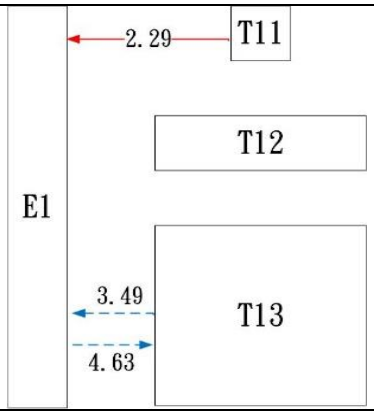
Eye fixations and saccades of different assessors

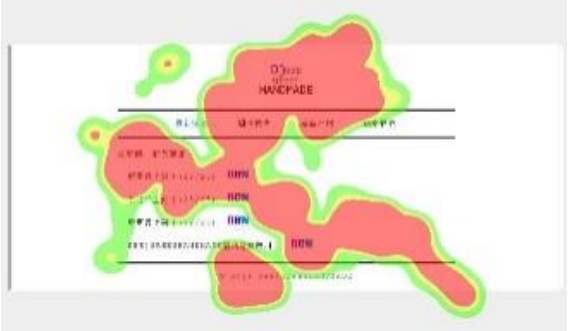
It was found that the regions which the reviewers giving higher scores gazed at were different from the regions which the reviewers giving lower scores fixated on. What they saw became a reference or reason for why they gave their peers' website high or low scores. This study explored whether the students giving high or low scores evaluated each website validly. The student designers provided the key regions of their design plan in each web page when they published the website. The key regions referred to the key points on the page which were expected to attract the viewers' attention. In this study, we compared the fixation heat map with the key points of the design plan.

For example, on the first page of website A, the students giving high scores frequently moved their eyes from the logo to the outside space area. From the saccade pattern, it was found that the student reviewers gave evaluation with lower scores when they gazed at the text at the bottom and looked around the space area, as shown in Table 8. The same behaviors of the student reviewers doing peer assessment and giving high and low scores were removed. Therefore, the different saccade behaviors of the student reviewers doing peer assessment with high and low scores were found. In addition, we also examined the eye hotspots of the student reviewers doing peer assessment with high and low scores. For example, in Table 8, it could be inferred that the ROI of T11 on website A may be the reason why points were deducted for the webpage, as other ROIs had similar percentages of fixation from the reviewers, but T11 had almost double the number of reviewers giving it lower scores. After interviewing the reviewers, they said that the logo was a little too small so they gazed at it more frequently at first.

On the contrary, when the web page design was poor, the careful reviewers gazed at different areas to find out where its key point was. For example, website B had a relatively poor web page design because its background could not clearly show the information content. Therefore, the students giving lower scores for website B had more patterns of saccades among the ROIs, as shown in Table 9. A comparison of the hotspots of the reviewers giving peer assessment with high and low scores showed that some students gave lower scores because they could not gaze at the numerous words for a long time because the design distracted them from the content. The eye-tracking provides triangulated evidence which confirms the scoring process.

Table 8. The different behaviors of the reviewers who gave peer assessments with high and low scores for an example page of website A

Different behaviors	An example of ROIs
	

Fixation partition of the reviewers who gave peer assessment with high scores	Fixation partition of the reviewers who gave peer assessment with low scores
	

T11	T12	T13	E1	T11	T12	T13	E1
3	7	7	6	8	10	10	6
30%	70%	70%	60%	57%	71%	71%	43%

Note. Red solid lines refer to the unique behavioral patterns of the students giving peer assessment with high scores. Blue dotted lines refer to the unique behavioral patterns of the students giving peer assessment with low scores.

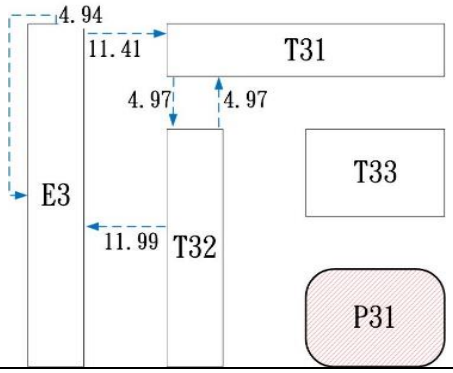
In sum, in terms of website A, seven of the 10 webpages had high scores with frequent and effective fixations on key points in the page during the peer assessment. From those behavior patterns, it could be concluded that the students giving high scores conducted an effective assessment of website B. With regard to the evaluation results given by the experts, the overall rating of website A was also high. The triangulated evidence confirmed that website A should be ranked top among the four websites.

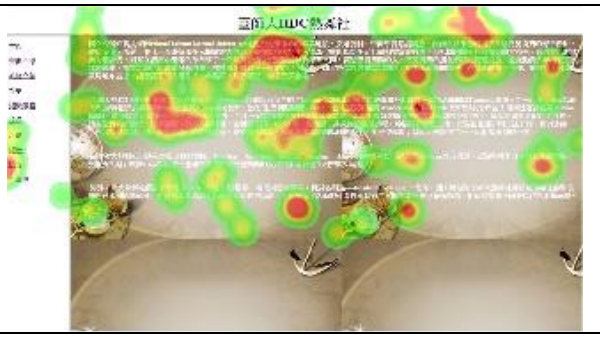
However, as for website B, only three of 10 webpages had high scores with frequent fixations on key points at the same time. Most students assessed website B with low scores, and those students concentrated on the design key points. In other words, the students giving low scores performed effective assessment, rather than those giving high scores. The overall ranking of website B was not so good according to the evaluation results of the experts. As for website C, it was worse, as seven webpages were given low scores with effective eye movement records. Website D had five web pages with high scores and the other five had low scores. The overall ranking of the four websites from high to low scores in the peer assessment was A, D, B, and C, which matched the ranking given by the experts.

On the other hand, the number of student reviewers gazing at the same ROIs as the experts was defined as concentration. This study selected the top 27% concentrated student reviewers (i.e., $N = 48 \times 0.27 = 13$), and the lowest 27% concentrated student reviewers (i.e., $N = 48 \times 0.27 = 13$) from all participants based on the fixation analysis. The results are displayed in Table 10. The experts assessed websites A and D as high scoring, while they appraised websites B and C as low scoring. The study explored how the high- and low-concentrated student reviewers scored the four websites. The evidence of the eye-tracking from the peer assessment process showed that website A had the best design of the four websites. The average of the scores given by the experts was not only the highest, but also the student reviewers all gave it high scores regardless of being low- or high-

concentrated. Website D had a similar situation. Conversely, websites B and C gained low scores no matter whether they were evaluated by high- or low- concentrated student reviewers.

Table 9. The different behaviors of the reviewers who gave peer assessments with high and low scores for an example page of website B

Different behaviors	An example of ROIs
	

Fixation partition of the reviewers who gave peer assessment with high scores	Fixation partition of the reviewers who gave peer assessment with low scores
	

T31	T32	T33	P31	E3	T31	T32	T33	P31	E3
5	6	9	5	4	7	5	10	5	6
38%	46%	69%	38%	31%	64%	45%	91%	45%	55%

Note. Red solid lines refer to the unique behavioral patterns of the students giving peer assessment with high scores. Blue dotted lines refer to the unique behavioral patterns of the students giving peer assessment with low scores.

Table 10. The cross-match between the heat maps and scores given by the experts and peers

Website	Results of expert assessment	Results of peer assessment	High concentration		Low concentration	
			N = 13	Consistency percentage	*N* = 13	Consistency percentage
A	High scores	High scores	13	100%	13	100%
		Low scores	0		0	
B	Low scores	High scores	0	100%	1	92.31%
		Low scores	13		12	
C	Low scores	High scores	1	92.31%	0	100%
		Low scores	12		13	
D	High scores	High scores	12	92.31%	13	100%
		Low scores	1		0	

Consequently, the low- and high-concentrated student reviewers all performed effective assessment because their evaluation achieved more than 90% similarity with the expert appraisal. The reason why the high-concentrated student reviewers were not able to achieve 100% effective assessment is inferred as follows. It could be said that high concentration is not the only necessary condition for highly effective peer review because the highly concentrated students may have focused on the areas which were not the key points of the page. When the students concentrated on the area which was not the main information place the designers wanted to deliver, they would not absorb the full message, meaning that some of their evaluation could not be effective or valid. Therefore, the heat maps alone could not ensure that the students' assessments were valid. The heat maps had to be triangulated and examined along with the quantitative data including the students' attitudes towards peer

assessment, the scores given by each student reviewer for every website, and the saccade patterns of the student reviewers.

Discussion

From the quantitative results, it was found that the consistency rate between the evaluation performed by the experts and that of the peers achieved more than 70%. This conforms to a previous study (Falchikov & Goldfinch, 2000) which indicated that providing student reviewers with comprehensive evaluation criteria or rubrics for assessing products could result in evaluation results like those of experts. A previous study has shown that understanding the assessment rubrics well could ensure that students complete the peer review fairly (Cheng & Warren, 1997). The investigation results of peer assessment attitude in the current study showed that the students possessed positive attitudes towards and understanding of peer assessment in this study. However, it should be noted that this study did not introduce the comments given by the students and experts. Future studies are encouraged to code such comments and to compare them with the eye movement clues.

After confirming that most students' assessment capabilities were close to those of the experts with the help of clear rubrics, it was then found that the assessment results were also affected by the diversity of the materials presented in the websites from the behavioral sequential analysis of the eye movements. In particular, poor web interface design caused a lower degree of consensus between the assessment of the experts and students. This study provides additional evidence and triangular validation for a previous study on the eye tracking technique (Iqbal & Mahmood, 2008), revealing that the assessment of teachers would be different from that of students when the students and teachers evaluate a disappointing creation (Iqbal & Mahmood, 2008).

The current study found that the grades given by the students tended to be higher than those given by the experts when the students were confronted with a design they were not familiar with. Another study indicated that the students gave higher scores in comparison with the teachers when the students faced an unfamiliar scoring item or target (Lindblom-ylänne et al., 2006). Based on this finding, it is suggested that teachers should avoid only exhibiting well-designed websites, but should increase the diversity of examples so that students could not only learn the advantages of well-designed webpages but could also think about how to improve poor designs. The appraisal capability of the students could be enhanced by such training.

In this study, it was also concluded what constitutes a better design for a web page. Website A was more concise and clearer than website D. From the heat map, it was found that the viewers were more concentrated when they evaluated Website A. Website A did not cause too much multimedia cognitive loading due to its clear design (Mayer & Moreno, 2003). A clearer design will reduce student reviewers' negative attitude toward peer assessment and promote their concentration. In addition, the design of the logo and menu was consistent on each page of Website A. Most of the students looked at the logo and menu before looking at the content. The design was close to the experience of the students with a visual learning style (Axelsson, 2012), and similar to another reading study which pointed out that the students had movement of the eye between the picture and its explanation when the picture and text were related (Mason et al., 2015).

It is also concluded what constitutes a poor design for a web page. In Websites B and C, it was not easy for the assessors to distinguish the background and words on most of the pages. Scholars have mentioned that web design would be evaluated poorly when the visibility of important content is unclear (Fleming & Koman, 1998) as the student reviewers would be easily distracted from the focal point. In order to check whether the designers' intentions are well presented on their web pages, the designers should point out the positions of the key points on each page. After comparing the key points with the viewers' fixations, the reviewers could be sure that they had the right focus when observing the website and that the designers did successfully deliver their subject. The design was unsuccessful when the student reviewers could not focus on the information which the designers mainly wanted to deliver. Finally, the students not only had to concentrate on the web page but also had to focus on the right place (Ariasi & Mason, 2011).

Conclusions

This study incorporated the evidence of eye movements with peer assessment and expert assessment to triangulate the validation in order to confirm whether the students who gave high scores or those who gave low scores conducted an effective peer assessment for a particular website, and to explain which webpage design attracted the viewers' attention to the right regions. We employed the eye movement technique to open the black

box of the peer review process. The analysis provides more scientific evidence to better understand the peer assessment process so that the reliability and validity of peer assessment could be enhanced.

The results indicated that the students giving higher and lower scores had different eye movements. When the higher scores given by the student reviewers were highly consistent with the evaluation of the experts, those who gave lower scores were relatively poor at peer assessment. Based on the eye fixation hotspot evidence, when the students were more concentrated on the peer assessment, their evaluated results were closer to those of the two experts. The eye fixation hotspots were the same as the key points planned by the student designers of the website which scored the highest. In other words, the eye-tracking behavioral records of peer assessment provided the students with additional important feedback for designing a website.

Finally, it should be pointed out that the major research limitation of this study is the limited number of times that peer assessment was performed. Future studies are encouraged to conduct more rounds of peer assessment so that the students have opportunities to repair their designs. The visual learning style was the control item in this study, so future studies could consider making use of different learning styles as a dependent variable.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology in the Republic of China under contract numbers MOST 105-2628-S-003-002-MY3. This work is also partly supported by the "Aim for the Top University Project" of the National Taiwan Normal University and the Ministry of Education, Taiwan, R.O.C. The authors would like to thank Prof. Hong-Fa Ho who provided the eye-tracking equipment and technical support.

References

Aeiad, E., & Meziane, F. (2015). An Adaptable and personalised e-learning system based on free web resources. In *International Conference on Applications of Natural Language to Information Systems* (pp. 293-299). Passau, Germany: Springer, Cham.

Ashton, S., & Davies, R. S. (2015). Using scaffolded rubrics to improve peer assessment in a MOOC writing course. *Distance Education, 36*(3), 312-334.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education, 27*(5), 427-441.

Barrios, V. M. G., Gütl, C., Preis, A. M., Andrews, K., Pivec, M., Mödritscher, F., & Trummer, C. (2004). AdELE: A Framework for adaptive e-learning through eye tracking. In *Proceedings of IKNOW* (pp. 609-616). Graz, Austria: Citeseer.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education, 24*(4), 413-426.

Buckley, P., & Doyle, E. (2017). Individualising gamification: An Investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. *Computers & Education, 106*, 43-55.

Calvi, C., Porta, M., & Sacchi, D. (2008, July). e5Learning, an e-learning environment based on eye tracking. In *The 18th IEEE International Conference on Advanced Learning Technologies (ICALT 2018)* (pp. 376–380). Mumbai, India: IIT Bombay.

Cassidy, S. (2006). Developing employability skills: Peer assessment in higher education. *Education+ Training, 48*(7), 508-517.

Chen, S. Y., & Yeh, C.-C. (2017). The Effects of cognitive styles on the use of hints in academic English: A Learning analytics approach. *Educational Technology & Society, 20*(2), 251-264.

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. *The Internet and Higher Education, 25*(2015), 78-84.

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*(2), 233-239.

Chiang, T. H., Yang, S. J., & Hwang, G. J. (2014). Students' online interactive patterns in augmented reality-based inquiry activities. *Computers & Education, 78*, 97-108.

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*(5), 629-643.

Claxton, C. S., & Murrell, P. H. (1987). *Learning styles: Implications for improving educational practices*. Washington, DC: Association for the Study of Higher Education. (ASHE-ERIC Higher Education Report No. 4)

Darban, M., & Polites, G. L. (2016). Do emotions matter in technology training? Exploring their effects on individual perceptions and willingness to learn. *Computers in Human Behavior, 62*, 644-657.

Dunn, K., & Dunn, R. (1987). Dispelling outmoded beliefs about student learning. *Educational Leadership, 44*(6), 55-63.

Fatahi, S., Moradi, H., & Kashani-Vahid, L. (2016). A Survey of personality and learning styles models applied in virtual environments with emphasis on e-learning environments. *Artificial Intelligence Review, 46*(3), 413-429.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering education, 78*(7), 674-681.

Fleming, N., & Baume, D. (2006). Learning styles again: VARKing up the right tree! *Educational Developments, 7*(4), 4-7.

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: A Multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning*, *31*(5), 435-449.

Godfroid, A., Housen, A., & Boers, F. (2010). A Procedure for testing the Noticing Hypothesis in the context of vocabulary acquisition. *Cognitive processing in second language acquisition*, 169-197.

Gütl, C., Pivec, M., Trummer, C., García-Barrios, V. M., Mödritscher, F., Pripfl, J., & Umgeher, M. (2005). ADELE (adaptive e-learning with eye-tracking): Theoretical background, system architecture and application scenarios. *European Journal of Open, Distance and E-Learning*, *8*(2).

Hawk, T. F., & Shah, A. J. (2007). Using learning style instruments to enhance student learning. *Decision Sciences Journal of Innovative Education, 5*(1), 1-19.

Hewig, J., Trippe, R. H., Hecht, H., Straube, T., & Miltner, W. H. (2008). Gender differences for specific body regions when looking at men and women. *Journal of Nonverbal Behavior, 32*(2), 67-78.

Hoffman, J. E., & Subramaniam, B. (1995). The Role of visual attention in saccadic eye movements. *Perception & psychophysics*, *57*(6), 787-795.

Hsu, C. K., Hwang, G. J., & Chan, C. K. (2014). An Automatic caption filtering and partial hiding approach to improving the English listening comprehension of EFL students. *Journal of Educational Technology & Society*, *17*(2), 270.

Hsu, C. K., Hwang, G. J., & Chang, C. K. (2010). Development of a reading material recommendation system based on a knowledge engineering approach. *Computers & Education*, *55*(1), 76-83.

Hsu, C. K., Hwang, G. J., Chang, Y. T., & Chang, C. K. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Educational Technology & Society, 16*(1), 403-414.

Huang, E. Y., Lin, S. W., & Huang, T. K. (2012). What type of learning style leads to online participation in the mixed-mode e-learning environment? A Study of software usage instruction. *Computers & Education*, *58*(1), 338-349.

Huang, Y. M., Liao, Y. W., Huang, S. H., & Chen, H. C. (2014). A Jigsaw-based cooperative learning approach to improve learning outcomes for mobile situated learning. *Educational Technology & Society, 17*(1), 128-140.

Hwang, G. J., & Chen, C. H. (2016). Influences of an inquiry-based ubiquitous gaming design on students' learning achievements, motivation, behavioral patterns, and tendency towards critical thinking and problem solving. *British Journal of Educational Technology, 48*(4), 950-971.

Hwang, G. J., Chu, H. C., & Yin, C. J. (2017). Objectives, methodologies and research issues of learning analytics. *Interactive Learning Environments, 25*(2), 143-146.

Hwang, G. J., Hsu, T. C., Lai, C. L., & Hsueh, C. J. (2017). Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers & Education, 106*, 26-42.

Hwang, G. J., Hung, P. H., Chen, N. S., & Liu, G. Z. (2014). Mindtool-assisted in-field learning (MAIL): An Advanced ubiquitous learning project in Taiwan. *Journal of Educational Technology & Society*, *17*(2), 4-16.

Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment & Evaluation in Higher Education, 29*(6), 751-768.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology, 8*(4), 441-480.

Kuo, Y. C., Chu, H. C., & Huang, C. H. (2015). A Learning style-based grouping collaborative learning approach to improve EFL students' performance in English courses. *Journal of Educational Technology & Society*, *18*(2), 284-298.

Lai, C. L., & Hwang, G. J. (2015). An Interactive peer-assessment criteria development approach to improving students' art design performance using handheld devices. *Computers & Education*, *85*, 149-159.

Lai, C. Y. (2016). Training nursing students' communication skills with online video peer assessment. *Computers & Education, 97*, 21-30.

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., Lee, M. H., Chiou, G. L., Liang, J. C., & Tsai, C. C. (2013). A Review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90-115. doi:10.1016/j.edurev.2013.10.001.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525-536.

Lin, S. J., Liu, Z. F., & Yuan, S. M. (2001). Web based peer assessment: Attitude and achievement. *Education, IEEE Transactions on, 44*(2). doi:10.1109/13.925865

Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An Eye-movement analysis. *Learning and instruction, 47*, 91-102.

Liu, Z. F., Lin, S. J., & Yuan, S. M. (2002). Alternatives to instructor assessment: A Case study of comparing self and peer assessment with instructor assessment under a networked innovative assessment procedures. *International Journal of Instructional Media, 29*(4), 395-404.

Lui, A., & Andrade, H. (2014). Student peer assessment. In R. Gunstone, (Ed.), *Encyclopedia of Science Education* (pp. 1-3). The Netherlands: Springer.

Lujan, H. L., & DiCarlo, S. E. (2006). First-year medical students prefer multiple learning styles. *Advances in Physiology Education, 30*(1), 13-16.

Luo, H., Robinson, A. C., & Park, J.-Y. (2014). Peer grading in a MOOC: Reliability, validity, and perceived effects. *Journal of Asynchronous Learning Networks, 18*(2), 1-14.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback, and L2 development. In P. Robinson (Ed.), *Individual differences in L2 learning* (pp. 181–210). Amsterdam, The Netherlands: Benjamins.

McLeay, F., & Wesson, D. (2014). Chinese versus UK marketing students' perceptions of peer feedback and peer assessment. *The International Journal of Management Education, 12*(2), 142-150.

Ng, E. M. (2016). Fostering pre-service teachers' self-regulated learning through self-and peer assessment of wiki projects. *Computers & Education, 98*(1), 180-191.

Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The Impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*(4), 195-203.

Pivec, M., Trummer, C., & Pripfl, J. (2006). Eye-tracking adaptable e-learning and content authoring support. *Informatica, 30*(1), 83-86.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372-422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506.

Russell, J., Van Horne, S., Ward, A. S., Bettis, E. A., & Gikonyo, J. (2017). Variability in students' evaluating processes in peer assessment with calibrated peer review. *Journal of Computer Assisted Learning, 33*(2), 178-190.

Schmidt, R. (1990). The role of consciousness in second language learning1. *Applied linguistics, 11*(2), 129-158.

Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. *Interlanguage pragmatics, 21*, 42.

Schmidt, R. (1995). Consciousness and foreign language learning: A Tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning, 9*, 1-63.

Schmidt, R., & Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Seng, M. G. J., & Hill, M. (2014). Using a dialogical approach to examine peer feedback during chemistry investigative task discussion. *Research in Science Education, 44*(5), 727-749.

Seyal, A. H., & Rahman, M. N. A. (2015). Understanding learning styles, attitudes and intentions in using e-learning system: Evidence from Brunei. *World Journal of Education, 5*(3), 61-72. doi:10.5430/wje.v5n3p61

Soloman, B. A., & Felder, R. M. (2005). *Index of learning styles questionnaire.* NC State University. Retrieved from http://www.engr.ncsu.edu/learningstyles/ilsweb.html

Sternberg, R. J., & Grigorenko, E. L. (1997). Are cognitive styles still in style? *American psychologist, 52*(7), 700-712.

Su, A., Yang, S. J., Hwang, W. Y., Huang, C. S., & Tern, M. Y. (2014). Investigating the role of computer-supported annotation in problem-solving-based teaching: An Empirical study of a Scratch programming pedagogy. *British Journal of Educational Technology, 45*(4), 647-665.

Tayem, Y. I., James, H., Al-Khaja, K. A., Razzak, R. L., Potu, B. K., & Sequeira, R. P. (2015). Medical students' perceptions of peer assessment in a problem-based learning curriculum. *Sultan Qaboos University medical journal, 15*(3), e376–e381.

Tenório, T., Bittencourt, I. I., Isotani, S., Pedro, A., & Ospina, P. (2016). A Gamified peer assessment model for on-line learning environments in a competitive context. *Computers in Human Behavior, 64*, 247-263.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research, 68*(3), 249-276.

Tsai, M. J., Huang, L. J., Hou, H. T., Hsu, C. Y., & Chiou, G. L. (2016). Visual behavior, flow and achievement in game-based learning. *Computers & Education, 98*, 115-129.

Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270-279.

Wang, S. M., Hou, H. T., & Wu, S. Y. (2017). Analyzing the knowledge construction and cognitive patterns of blog-based instructional activities using four frequent interactive strategies (problem solving, peer assessment, role playing and peer tutoring): A Preliminary study. *Educational Technology Research and Development, 65*(2), 301-323.

Wang, Y., Liang, Y., Liu, L., & Liu, Y. (2016). A Multi-peer assessment platform for programming language learning: Considering group non-consensus and personal radicalness. *Interactive Learning Environments, 24*(8), 2011-2031.

Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*(1), 27-44.

White, E. (2009). Student perspectives of peer assessment for learning in a public speaking course. *Asian EFL Journal, 33*(1), 1-36.

Xie, K. (2013). What do the numbers say? The Influence of motivation and peer feedback on students' behaviour in online discussions. *British Journal of Educational Technology, 44*(2), 288-301.

Yang, F. Y., Huang, R. T., & Tsai, I. J. (2016). The Effects of epistemic beliefs in science and gender difference on university students'science-text reading: an eye-tracking study. *International Journal of Science & Mathematics Education, 14*(3), 473-498.

Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., & Ogata, H. (2015). Identifying and analyzing the learning behaviors of students using e-books. In *Proceedings of 1st Workshop on e-Book-based Educational Big Data for Enhancing Teaching and Learning on ICCE 2015* (pp. 617-626). Hangzhou, China: ICCE 2014 Organizing Committee.

Yu, F. Y., & Wu, C. P. (2013). Predictive effects of online peer feedback types on performance quality. *Journal of Educational Technology & Society, 16*(1), 332-341.