Computers & Education 116 (2018) 225–236

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Exploring effects of discussion on visual attention, learning performance, and perceptions of students learning with STR-support

Wu-Yuin Hwang[a], Yung-Hui Li[b], Rustam Shadiev[c,*]

[a] Graduate Institute of Network Learning Technology, National Central University, No.300, Jhongda Rd., Jhongli 32001, Taiwan
[b] Department of Computer Science and Information Engineering, National Central University, No.300, Jhongda Rd., Jhongli 32001, Taiwan
[c] School of Education Science, Nanjing Normal University, No. 122, Ninghai Road, Nanjing 210097, Jiangsu Province, China

ARTICLE INFO

Keywords:
Cooperative/collaborative learning
Improving classroom teaching
Pedagogical issues

ABSTRACT

In this study, we aimed to explore effects of discussion on visual attention, learning performance, and perceptions of non-native English speaking students attending lectures in English in learning environments supported by Speech-to-Text Recognition (STR) system. One experiment was carried out with 60 students who were assigned into the control (n = 30) and experimental (n = 30) groups. Students in the control group attended lectures without discussing lectures content whereas students in the experimental group attended lectures and discussed lectures content. Our result showed that students in the experimental group had higher fixation time during two lectures compared to their counterparts in the control group. We also found that students in the experimental group had higher fixation count during the first lecture compared to the control group but there was no difference between the two groups in fixation count during the second lecture. Our results demonstrated that the experimental group outperformed the control group on both tests. In addition, most students in the two groups had high perceptions towards usefulness of STR-text for learning; however, no differences were found between the two groups in their perceptions towards STR-text. Based on our results, we suggest that learning activities, such as student discussion, need to be introduced during lectures in order to stimulate active learning, which in turn, enhances students' learning and comprehension of lecturing content. Discussion may encourage and facilitate students to take more active role in their learning.

1. Introduction

1.1. Speech-to-text recognition

Speech-to-Text recognition (STR) technology has received much attention recently in the research literature. STR synchronously generates text from a lecturer's speech and displays it for students on a whiteboard or computer screens. Recent evidence suggests that applications of STR technology are beneficial for learning (Huang, Liu, Shadiev, Shen, & Hwang, 2015). For example, during lectures, STR-text helps students overcome difficulties in reading, writing, and spelling (Nisbet, Wilson, & Aitken, 2005), assists participation of deaf students (Leitch, 2008; Wald & Bain, 2008), facilitates comprehension of learning content (Ranchal et al., 2013),

* Corresponding author.
E-mail address: rustamsh@gmail.com (R. Shadiev).

http://dx.doi.org/10.1016/j.compedu.2017.09.015
Received 29 November 2016; Received in revised form 24 September 2017; Accepted 29 September 2017
Available online 30 September 2017

especially of non-native speakers (Ryba, McIvor, Shakir, & Paez, 2006; Shadiev, Hwang, Huang, & Liu, 2016; Wald & Bain, 2008) and students in online synchronous learning environments (Hwang, Shadiev, Kuo, & Chen, 2012; Kuo, Shadiev, Hwang, & Chen, 2012).

1.2. Multimedia related theories

Mayer (2010) argued that learning is an active process and for meaningful learning to occur a learner needs to be engaged in appropriate cognitive processing during learning such as selecting of relevant learning content, organizing it into a coherent cognitive representation, and integrating it with prior knowledge activated from long-term memory. In order to understand the usefulness of STR-text for learning, we need to refer to a cognitive theory of multimedia learning which explains how a learner learns with content presented in different modalities (Mayer & Moreno, 2003). According to this theory, visual and verbal information is received and decoded through different processing systems. The visual channel takes input initially from the eyes and ultimately produces pictorial representations whereas the verbal channel takes input from the ears and decodes them into verbal representations. When listening to speech, a learner pays attention to the auditory message, parses speech, segments it into words that are kept in verbal working memory, and then transforms words into verbal mental representations. After this process, the connections among words are mentally built to organize them into cause-and-effect chains. Similarly, a learner pays attention to visual messages, selects images, and retains them in visual working memory. A learner also mentally builds connections that organize the images into a cause-and-effect chain. Finally, the visual mental model, verbal mental model, and prior knowledge from long-term memory are merged through constructing referential connections between them (Mayer, 2009). Thus, presenting verbal and visual information simultaneously helps process learning content better. During lecture, when the instructor delivers lecturing content and STR generates text from the speech, which is displayed for students simultaneously, students receive information in two modalities - verbal (i.e. speech of the instructor) and visual (i.e. STR-text). Learning information presented in two modalities enables students to attain a better understanding of lecture content.

According to cognitive load theory (Paas, Renkl, & Sweller, 2003), students working memory has limited cognitive capacity to accommodate demands imposed by learning tasks. When students attend a lecture in a foreign language, their working memory is at risk to be cognitively overloaded. The reason is because students need to receive and retain lecturing information in working memory and integrate it with what follows, all the while continually adjusting their understanding with prior knowledge (Chen & Chang, 2009). When students' cognitive load exceeds the limit of cognitive capacity, it may negatively affect learning performance (Mayer & Moreno, 2003; Paas et al., 2003). In our study, students attended lectures in English which was their foreign language. In order to facilitate knowledge construction and to reduce cognitive capacity overload during lectures in a foreign language, we provided students with STR-text. So students could refer to STR-text during lecture, for example, to textually confirm what is being said in lecture.

However, it is suggested that the same information presented in two modalities (i.e. auditory and visual) increases cognitive load as presented information becomes redundant and gives rise to a split-attention effect (Mayer, Lee, & Peebles, 2014). That is, it is possible that presenting STR-text during lectures may result in the redundancy effect since the same information will be presented in verbal (i.e. a lecturer's speech) and visual (STR-text) forms simultaneously. On the other hand, Clark and Mayer (2011) claimed that there are exceptions to the redundancy principle; that is, redundant visual text can be acceptable and even necessary along with the verbal information, in specific situations. For example, a situation when the verbal information is difficult to understand, which commonly happens when students attend lectures in a foreign language. Another consideration, which relates to multimedia learning and cognitive load, is the expertise reversal principle (Kalyuga, 2014) and it also needs to be seriously considered. According to this principle, the effectiveness of instructional techniques on students with differing levels of prior knowledge can be reverse. That is, instructional techniques that are highly effective with novice learners may not be effective when used with more knowledgeable learners. For example, students with low language ability and lack of prior knowledge need learning information presented in visual and verbal forms to comprehend it better. On the other hand, students with high language ability and prior knowledge need information to be presented in one modality only. Multimodal information can be redundant for such students as it requires additional cognitive resources to process it. So in this study, we attempted to explore whether STR-text can be useful for students during lectures in a foreign language and for what kind of students, i.e. low ability or high ability.

1.3. Discussion

Discussion, when students talk about learning material to and exchange their ideas with other students, plays a valuable role in lecture or seminar. Noroozi, Weinberger, Biemans, Mulder, and Chizari (2013), Noroozi, Weinberger, Biemans, Mulder, and Chizari (2012), and Stegmann, Wecker, Weinberger, and Fischer (2012) suggested that discussion is a learning activity that engages students in small groups in argumentation, critical thinking, elaboration, and reasoning processes. Discussion facilitates individual knowledge acquisition and collaborative elaboration. In addition, discussion enables students to build up a shared understanding of learning content. There are several advantages of discussion, which were highlighted in the literature. Grzega and Schoner (2008) argued that discussion encourages and stimulates students to take a more active role in their learning. Noroozi et al. (2013, 2012) and Stegmann et al. (2012) suggested that students share their knowledge with learning partners and it leads to greater cognitive effort. Students not only confront each other with different viewpoints but contribute diverse information to discussion as well. Students articulate a variety of opinions and points of view and acquire knowledge and insight from diverse points of view (Crippena & Earl, 2007). Discussion offers students opportunities to test their ideas and opinions against those of their peers (Aleven & Koedinger, 2002). In addition, discussion helps students understand and apply what they have learned (Hubscher, 2010). In addition, during discussion, students have a chance to explain concepts they learned to other students. This process leads to the construction of additional knowledge by the explaining student and his/her self-reflection on own learning progress (Grzega & Schoner, 2008). The explainer may identify missing information, point out inconsistencies, and needs for further clarification while explaining to the receiver whereas the receiver may also identify similar issues while being explained. Crippena and Earl (2007) suggested that in order to be prepared for discussion, students learning need to go beyond mere reviewing of or listening to learning content. Noroozi et al. (2013) argued that students need to allocate a considerable part of their time and cognitive capacity to prepare for discussion. That is, apart from reviewing/listening to learning content, students need to think analytically about it, to relate it to what they already know, and to use that knowledge for constructing new knowledge, solving new problems, and addressing new issues (Hubscher, 2010). This is because students not only need to understand learning content but also be able to explain it to other students. As a result, advanced cognitive processes are activated. In order to facilitate discussion and make it more efficient, Noroozi et al. (2013) suggested providing students with textual representations, e.g. scripts. They argued that textual representations are useful for individual acquisition of knowledge and joint knowledge construction. In this study, students used STR-text during discussion. We assumed that STR-text, as textual representations, can help students to comprehend learning content as well as facilitate their discussion better. For example, using STR-text, students may prepare for discussion and then explain learning content to other students. In addition, students may ask questions and challenge a partner when he/she explains learning content during discussions. Although discussion is beneficial and important for learning, review of STR-related research showed that most studies in the field have only focused on individual learning of students with or without STR-text. Previous studies have not examined how students' discussion of lecture content with STR support can be beneficial for learning. That is, it is not clear whether discussion will help or hinder learning of students in STR-supported environment.

1.4. Visual attention

Recently, researchers have shown an increasing interest in the eye-tracking method to study visual attention because it has a capacity to record and provide rich information about learning behaviors and learners' cognitive processes (Kruger & Steyn, 2014; Liu, 2014). According to the eye-mind hypothesis, the eye-tracking method can reveal the temporal change of visual attention that may further inform how learners approach and process learning information (Just & Carpenter, 1980). Several important research variables were considered in previous eye-tracking research. Fixations are the main measurements and indicate moments when the eyes are relatively stationary, taking in or encoding information (Jacob & Karn, 2003; Poole & Ball, 2006). Learning information processing and amount of attention are reflected by eye fixation duration (Tsai, Hou, Lai, Liu, & Yang, 2012). The longer the information is fixated, the deeper it is processed (Poole & Ball, 2006). Thus, Rayner (2009) claimed that the fixation duration indicates the period of time needed to acquire new information. Kang (2014) suggested that the fixation count is a measure related to the level of a learner's interest on a certain area because relevant information attracts more attention than irrelevant one. According to Rayner (2009) learners tend to have more fixations on more important information. Therefore, eye fixation count reveals the importance of learning material. Several multimedia learning studies, which employed the eye-tracking method, have been published recently. Students learned vocabulary in multimedia environments and their visual attention was examined by Liu (2014). Non-native English speaking students learned from computer-based lessons presented in English and researchers explored what multimedia content is more helpful, i.e. video or on-screen captions (Mayer et al., 2014). Kang (2014) investigated online reading patterns and comprehension of mother-tongued and non-native readers. Kruger and Steyn (2014) analyzed how subtitle reading influences non-native English learners to learn from lectures delivered in English. In our study, we employed the eye-tracking method to reveal how much attention a student pays on STR-text while attending a lecture. Such knowledge is very essential for the instructor in designing the curriculum and adopting STR technology. In addition, to the best of our knowledge, researchers have not dealt with visual attention of students on STR-text when they learn individually or in collaboration with other students and differences in attention of students learning under different conditions.

There is a relationship among research variables of our study and this is why they are being measured. For example, attention is very important for learning and it affects student performance (Liu, 2014; Mayer et al., 2014). Due to attention, a student selects some important information and processes it, so that learning takes place (Poole & Ball, 2006; Rayner, 2009). If students pay more attention to learning information, they can process it deeper and thus comprehend better (Tsai et al., 2012). As a result, their performance will be better too (Liu, 2014; Tsai et al., 2012). On the other hand, if students do not pay attention to learning information, they won't be able to learn and perform well (Kang, 2014; Kruger & Steyn, 2014). Having positive perceptions towards a treatment is also important, both for attention and learning. Technology Acceptance Model theory was proposed by Davis (1989) to model how users come to accept and use a technology. According to this theory, several factors contribute to users' decision about how and when they will use a technology. For example, perceived usefulness and behavioral intention are such factors; the former represents the degree to which a user believes that using a technology would enhance his or her performance and the later measures the likelihood of a user employing a technology. If students perceive that STR-text is useful for their learning, they will readily use it to enhance comprehension of learning content. That is, students will pay more attention on STR-text during lectures, comprehend learning content better and as a result, have higher learning performance. On the other hand, if students perceive that STR-text is not useful for learning, they will not pay too much attention on it.

In this study, we aimed to explore the effects of discussion on visual attention, learning performance and perceptions of non-native English speaking students attending lectures in English in learning environments supported by Speech-to-Text Recognition (STR) system. Therefore, the following research questions were addressed:

Table 1
Participant demographic profiles.

Demographic characteristics	Control group		Experimental group	
	Frequency	Percentage	Frequency	Percentage
Gender				
Female	4	13.33	9	30.00
Male	26	86.67	21	70.00
Age (years old)				
18-22	27	90.00	30	100.00
> 24	3	10.00	0	0.00
Department				
Applied science	30	100.00	30	100.00
Study for the degree				
Undergraduate	23	76.67	30	100.00
Graduate	7	23.33	0	0.00
EFL proficiency level				
High	10	33.33	10	33.33
Low	10	33.33	10	33.33

(1) How different students' attention is when they learn lectures content with vs. without discussing it?
(2) Do students who attended a lecture and discuss its content perform differently than students who attended a lecture without discussing its content?
(3) What are perceptions of students across two conditions towards STR-texts?

2. Methods

2.1. Participants

Sixty university students majored in applied science participated in this study. All students were Chinese native speakers and English was their foreign language. We assigned students to the control (n = 30) and experimental (n = 30) groups. There was no significant difference in student distribution across two groups with respect to gender, age, academic degree, high/low English as a foreign language (EFL) proficiency level, and EFL proficiency level in general (see Table 3) ($p > 0.05$). This approach allowed us to ensure that the difference between the control and experimental groups in terms of research variables, which we measure in our study, is not due to the difference in student demographics. Background details of our students are shown in Table 1. Most students in the two groups were undergraduates, males, aged between 18 and 22.

2.2. Experimental procedure

The experimental procedure is presented in Fig. 1. In the beginning, we collected students' demographic information. After that, the students attended two lectures given in English. The first lecture, titled “Photography,” was at the intermediate level and the second one, “Tainan's great food,” was advanced. The first lecture was about two friends taking pictures with either a large professional camera or a small digital one whereas the second lecture was about three friends trying famous local food in Tainan city. The lectures were selected following general recommendations of earlier STR-related research (Shadiev, Huang, & Hwang, 2017; Shadiev, Wu, & Huang, 2017). These lectures were specifically developed for EFL learners and cover the intermediate and advanced EFL levels. In addition, lectures were not on a specific topic but a general one so no prior knowledge was required. These lectures were not a part of participants' coursework or curriculum. Both lectures were delivered to the students through computer screens. A screenshot of the second lecture is shown as Fig. 2. Students in both groups could see (a) the instructor, who lectured lectures content, (b) Power Point slides of the lecture, and (c) STR-generated text of the lecture. All the three media were synchronized. That is, when

Table 2
The results of fixation time and fixation count on STR-texts and *t*-test.

	Control group		Experimental group		t	p	η2
	Mean	SD	Mean	SD			
Lecture 1							
Fixation Time	30269.40	24632.91	42125.80	11338.41	−2.395	0.021	0.09
Fixation Count	90.13	90.74	144.40	48.52	−2.889	0.006	0.13
Lecture 2							
Fixation Time	37439.17	34668.98	52947.20	15349.57	−2.240	0.031	0.08
Fixation Count	246.93	672.93	161.73	39.25	0.692	0.494	0.01

Table 3
The results of the post-tests and analysis of covariance.

Test/Item	Control group (n = 30)		Experimental group (n = 30)		F	p	η^2
	Mean	SD	Mean	SD			
Covariate							
EFL proficiency	12.13	2.03	11.87	1.99	–	–	–
Post-test 1							
Item 1	2.03	1.25	2.70	0.84	6.836	0.011	0.11
Item 2	0.53	0.73	1.13	0.78	9.581	0.003	0.14
Item 3	0.80	0.71	1.27	0.58	7.725	0.007	0.12
Item 4	1.97	1.16	2.10	1.21	0.160	0.691	0.00
Total	**5.33**	**2.77**	**7.20**	**2.28**	**8.374**	**0.005**	**0.13**
Post-test 2							
Item 1	3.40	1.22	3.37	0.85	0.000	0.998	0.00
Item 2	0.60	0.50	0.90	0.30	9.047	0.004	0.14
Item 3	0.13	0.35	0.53	0.63	9.067	0.004	0.14
Item 4	1.73	1.17	2.20	1.32	1.924	0.171	0.03
Total	**5.87**	**2.08**	**7.00**	**1.84**	**4.983**	**0.030**	**0.08**

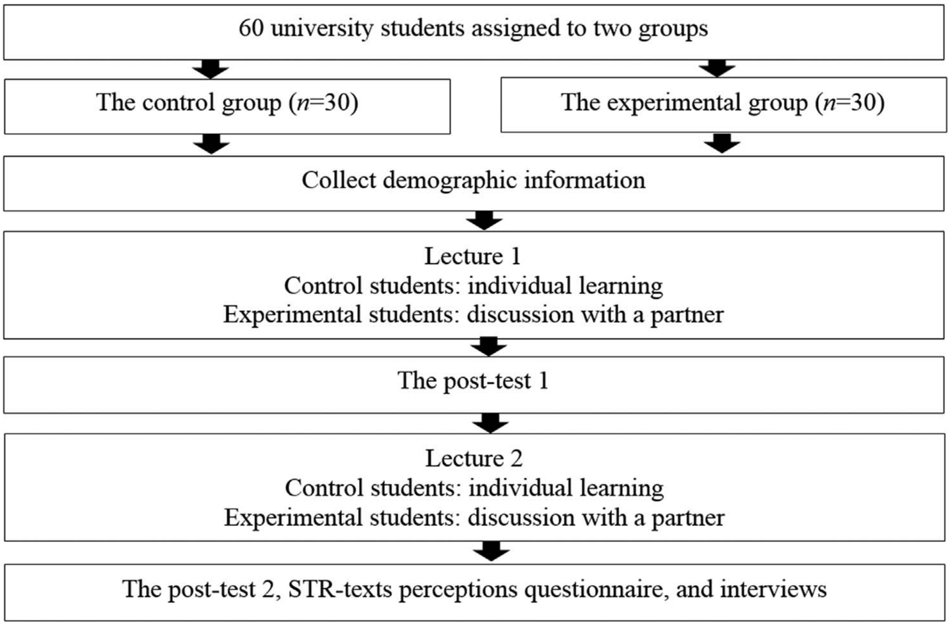

Fig. 1. Experimental procedure.

the instructor was lecturing about specific part of a lecture, a Power Point slide related to that part was shown. In addition, texts-generated by STR from the instructor's speech was shown synchronously to students. The main reason for providing students with STR-text was to aid their learning of lecturing content presented in EFL. English was merely the language of instruction during our lectures but not the subject the students were learning. The control group studied lecturing content individually. On the other hand, we extended individual learning of students in the experimental group into collaborative one. That is, the students were divided into pairs; they studied lecturing content individually and then discussed it for about 10–15 min with their partners. Students discussed content related to a lecture, particularly its parts which were difficult to comprehend due to complexity or novelty of content and language proficiency. There were no any observers, discussion leaders, or facilitators during discussion. Students were actively engaged in discussion because they aimed to enhance their comprehension of lecturing content and discussion was limited to pair discussion only; so there was no development of discussion to a class-wide one. Our choice for choosing discussion out of many other learning activities was made based on its advantages for learning emphasized in earlier studies (Aleven & Koedinger, 2002; Hubscher, 2010; Grzega & Schoner, 2008; Noroozi et al., 2013; Stegmann et al., 2012). For example, discussion encourages students to take a more active role in their learning, share their knowledge with other students, articulate a variety of opinions and points of view as well as acquire knowledge and insight from diverse points of view. Therefore, the difference between two conditions was in learning method: individual learning (the control group) versus individual learning extended with discussion (the experimental group). The post-test was carried out after each lecture. Finally, the questionnaire survey was filled out, and one-on-one semi structured interviews were conducted with the experimental group students at the end of the experiment.

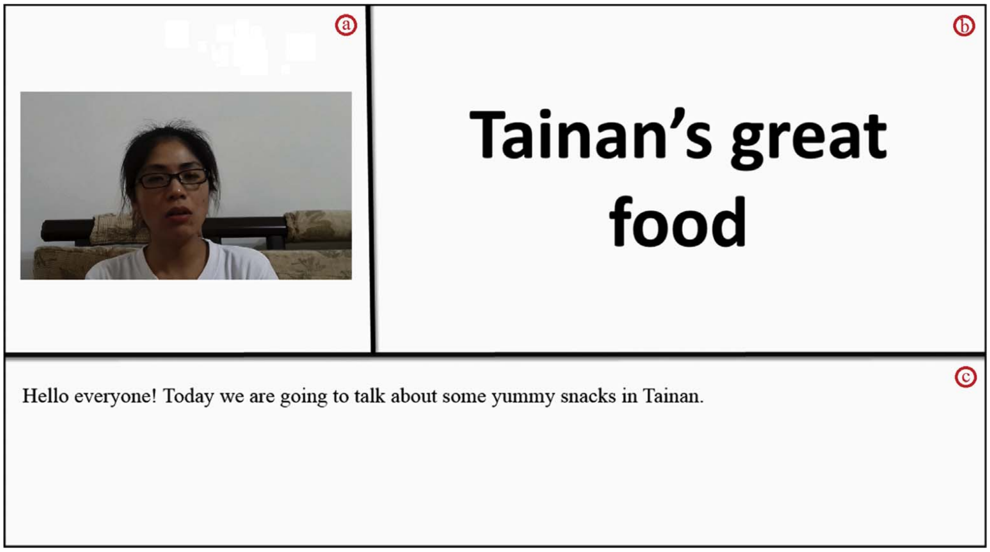

Fig. 2. A screenshot of the second lecture.

2.3. Research tools

2.3.1. EFL proficiency and learning performance evaluation

The content of the two lectures was general and unique. For example, the instructor introduced two characters in the lecture related to photography, and mentioned that the first character has a small digital camera whereas the second character has a large professional one. This is why it was impossible to assess student prior knowledge of this type of information. So we used students' current scores on EFL proficiency test of university examination as pre-test scores. We designed a post-test to measure students' learning performance. The post-test included four items; they related to the lectures content and were scored on a 15-point scale. The first item related to lecture information recognition; it consisted of 5 questions and students could get one point for each question answered correctly (maximum 5 points). The second and third items related to lecture information recall and consisted of one question each; students could get maximum 2 points for the second item and 3 points for the third item. The fourth item with one question measured student understanding of the lecture with maximum of 5 points. Three sample items of the posttest measuring information recognition, information recall, and understanding of the second lecture respectively are provided as following:

(1) In India
- A father reads the horoscope.
- A father should have a good education.
- A father meets relatives to discuss the marriage.
- A father finds his daughter a husband.

(2) What is the “business” part of a marriage agreement?
(3) Please summarize the content of the lecture.

Answers to post-test items were scored by three raters and major differences in assessment were resolved through discussion. Inter-rater reliability of the post-test was evaluated using Cohen's kappa and results exceeded 0.90, indicating high reliability.

2.3.2. Measuring perceptions

The questionnaire survey was administered to the students in order to investigate their perceptions toward usefulness of STR-text for learning and behavioral intentions to use STR-text in the future. The questionnaire was designed based on guidelines provided in related research (Hwang et al., 2012; Kuo et al., 2012) and included the following three dimensions: (1) Perceived usefulness of STR-text for learning - the degree to which a student believes that using STR-text during lectures would enhance his/her learning performance; (2) Behavioral intention to use STR-text - the degree to which a student intends to use STR-text for learning in the future. In addition, the following dimension was included in the questionnaire for the experimental group: (3) Perceived usefulness of STR-text during discussion - the degree to which a student believes that using STR-text during discussion would enhance his/her learning performance. Sixty valid answer sheets to the questionnaire were obtained from the students. This study ranked responses to the questionnaire using a five-point Likert scale, anchoring by the end-points “strongly disagree” (1) and “strongly agree” (5). We used Cronbach α to implement reliability analyses. The values were 0.885 for the first, 0.900 for the second, and 0.803 for the third dimensions demonstrating high reliability of the items.

2.3.3. Interviews

We conducted one-on-one semi-structured interviews with the students to explore the effectiveness of using STR-text for learning. The interviews contained open-ended questions in which the students were asked about: a) their experience using STR-text, and b) their opinions about the impact of STR-text for learning. Each interview took approximately 30 min. We audio-recorded all of the interviews and then fully transcribed for analysis. We highlighted and coded the text segments that met the criteria for providing the best research information according to an open coding method of analysis described by Strauss and Corbin (1998). After that, we sorted the codes into six categories: (1) Learning difficulties during lectures in a foreign language; (2) Learning behavior to use STR-texts during lecture with discussion; (3) Learning behavior to use STR-texts during lecture without discussion; (4) Usefulness of STR-texts for learning during lecture with discussion; (5) Usefulness of STR-texts for learning during lecture without discussion; and (6) Learning preference for media during lectures in a foreign language. Established categories produced a framework to illustrate findings relevant to the research questions.

2.3.4. Visual attention measurements

We employed eye-tracking technique to measure visual attention of participants in lectures. The eye-tracking technique composed of three main parts: (1) two monitors, one to present learning content to participants and the other to monitor the whole experiment; (2) the Microsoft LifeCame HD-6000 camera with 1280 × 720 pixels resolution to record the video around the ocular region (LifeCam HD-6000, 2017); (3) Ogama software to record, analyze and visualize gaze in screen based environments (OGAMA, 2017). The experimental stimulus was presented to participants on the monitor via OGAMA software and all data related to visual attention was recorded via Microsoft LifeCame HD-6000 camera. The data was then analyzed using OGAMA software.

The three areas of interests were defined for the learning material: (1) video of the instructor, (2) slides of the lecture, and (3) STR-text. However, we only focused on “STR-text” area of interest. The main reason for doing so is because we aimed to explore visual attention of the students on STR-text and explore the difference in attention between the two groups. Our eye-tracking system recorded the data related to visual attention. The following research variables for the eye-movement measures were used: (1) fixation time (FT) – the average time spent looking at an area of interest; it represents how deeply a student is involved to learn content; and (2) fixation count (FC) – the number of fixation points in an area; the fixation count for an area increases as a student becomes more interested in that area; therefore, fixation count can represent how important/interesting an area is to a student.

2.3.5. Speech-to-text recognition technology

We employed Windows® Speech Recognition using the Microsoft® Operating System as STR tool for this study (Shadiev & Huang, 2016). This tool is available for users of Microsoft Windows® Operating System at no additional cost (Huang, Shadiev, & Hwang, 2016). It is similar to a variety of commercial and open-source products in terms of performance and ease-of-use (Way, Kheir, & Bevilacqua, 2008, pp. 72–77).

2.4. Statistical analyses

First, we explored the difference in fixation time and fixation count on STR-text between the two groups using independent samples *t*-test. We examined the difference in post-test scores of the two groups, controlling for scores of the pre-test. For this examination we used an analysis of covariance. We employed independent samples *t*-test to compare perceptions towards STR-text of students in the two groups. A 5% significance level ($p < 0.05$) was used as a guideline for determining significant effects of variables for our statistical tests.

3. Results

3.1. Visual attention on STR-text

The means and standard deviations of the visual attention variables and the results of the statistical analysis are presented in Table 2. There was a statistically significant difference between the control ($M = 30269.40$; $SD = 24632.91$) and experimental ($M = 42125.80$; $SD = 11338.41$) groups with regard to fixation time during Lecture 1, $t = -2.395$, $p < 0.05$, $\eta^2 = 0.09$. That is, the fixation time value for the experimental group was higher than that for the control group. The difference in fixation count between the control ($M = 90.13$; $SD = 90.74$) and experimental ($M = 144.40$; $SD = 48.52$) groups during Lecture 1 was also significantly different, $t = -2.889$, $p < 0.05$, $\eta^2 = 0.13$. The control group had lower fixation count value compared to the experimental group. In addition, we found a significant difference between the control ($M = 37439.17$; $SD = 34668.98$) and experimental ($M = 52947.20$; $SD = 15349.57$) groups in fixation time during Lecture 2, $t = -2.240$, $p < 0.05$, $\eta^2 = 0.08$. That is, the fixation time value for the experimental group was higher. However, there was no difference in fixation count between two groups during Lecture 2, $t = 0.692$, $p > 0.05$, $\eta^2 = 0.01$. Big standard deviation value of fixation count for the control group ($SD = 672.93$) suggests that there could be a few outliers in the control group. We found that three students in the control group had unusually high values of fixation count (e.g. 3746) after careful examination of the data. We performed another statistical analysis but excluding the data of these three students from it. New results showed a significant difference between the control ($M = 103.70$, $SD = 101.34$) and experimental ($M = 161.73$, $SD = 39.25$) groups with regard to fixation count during Lecture 2, $t = -2.793$, $p < 0.05$, $\eta^2 = 0.12$.

Table 4
The results of the questionnaire and *t*-test.

Item	Control group		Experimental group		t	p	η^2
	M	SD	M	SD			
Perceived usefulness of STR-texts for learning	4.15	0.90	4.41	0.51	1.377	0.174	0.03
Behavioral intention	4.14	0.94	4.05	0.79	0.401	0.690	0.00
Perceived usefulness of STR-texts for discussion	NA	NA	4.39	0.66	NA	NA	NA

3.2. Learning performance

The means and standard deviations of the assessments and the results of the statistical analysis are summarized in Table 3. There was no significant difference between the pre-test scores of the students in the two groups, $t = 0.513$, $p = 0.610$, $\eta^2 = 0.00$. We found a significant difference in the post-test scores between the control and experimental groups with respect to the first ($F(1, 57) = 8.374$, $p < 0.05$, $\eta^2 = 0.13$) and second ($F(1, 57) = 4.983$, $p < 0.05$, $\eta^2 = 0.08$) lectures after controlling for the pre-test scores. That is, the experimental group performed better on both tests compared to the control group.

We further compared how differently students in the two groups performed on post-test items. Our results showed that students in the experimental group outperformed their counterparts on items 1, 2, and 3 of the post-test after Lecture 1 and items 2 and 3 of the post-test after Lecture 2, $p < 0.05$. There was no significant difference in performance of students across two groups on item 4 of the post-test after Lecture 1 and item 1 and 4 after Lecture 2, $p > 0.05$.

In order to explore what independent variable can significantly predict dependent variable, we carried out an ordinary least squares regression. We included post-test scores as a dependent variable and the experiment (having discussion), fixation time, fixation count, and EFL proficiency as independent variables in the model. The results of the regression indicated that for Lecture 1, the independent variables explained 21.9% of the variance ($R^2 = 0.219$, $F(3,26) = 2.431$, $p > 0.05$). It was also found that the independent variables do not significantly predict dependent variable ($p > 0.05$). For Lecture 2, the independent variables explained 15.5% of the variance, $R^2 = 0.155$, $F(3,26) = 1.593$, $p > 0.05$. In addition, we found that the independent variables do not significantly predict dependent variable ($p > 0.05$).

3.3. Perceptions towards STR-text

The questionnaire results are reported in Table 4. According to the table, most students in the control ($M = 4.15$; $SD = 0.90$) and experimental ($M = 4.41$; $SD = 0.51$) groups had high perceptions towards usefulness STR-text for learning. In addition, most students in the experimental group ($M = 4.39$; $SD = 0.66$) highly valued usefulness of STR-text for discussion. Finally, most students in the control ($M = 4.14$; $SD = 0.94$) and experimental ($M = 4.05$; $SD = 0.79$) groups intended to use STR-text for learning in the future. The results of statistical analysis revealed no significant difference in perceptions between the two groups for the first ($t = 1.377$, $p > 0.05$, $\eta^2 = 0.03$) and second ($t = 0.401$, $p > 0.05$, $\eta^2 = 0.00$) dimensions.

4. Discussion

Our statistical results showed that students in the experimental group had longer fixation time during two lectures and more fixation counts during the second lecture compared to the control group. As for fixation count during the second lecture, we did not find any differences between the two groups. Research on visual attention suggests that visual attention data, such as eye fixation, reflects how students process learning information and amount of their attention (Tsai et al., 2012). Higher values of eye fixation data indicates that learning information is important and processed deeper (Kang, 2014; Poole & Ball, 2006; Rayner, 2009). The fixation duration indicates the period of time needed to acquire new information (i.e. the longer the information is fixated, the deeper it is processed) whereas fixation count represents the interest level of a learner on a certain area of learning material (i.e. learners tend to fixate longer on more important information). Based on our results, we may conclude that the experimental group processed learning information deeper compared to the control group. From the interviews, we found that the reason was due to discussion that the experimental group had during lectures and for which they had to be prepared in advance. Students in the experimental group stated that to be prepared for discussion, they not only merely reviewed STR-texts and/or listened to the instructor but tried to understand it in order to be able to explain it to other students. They learned lectures content by reading STR-texts and/or listened to the instructor thoroughly and tried their best to comprehend it deeper and their advanced cognitive processes were activated (Crippena & Earl, 2007; Hubscher, 2010). This result was confirmed by questionnaire survey results (please refer to our discussion of questionnaire survey results later in this section).

Based on our results related to fixation count, we may conclude that the experimental group found STR-text as more important media compared to the control group. Our results may suggest that STR-text was more important for the experimental group and as a result, the experimental group paid more attention to STR-text to be well prepared for discussion. On the other hand, the control group didn't rely on STR-text as much as students in the experimental group did. It is suggested that fixation count represents the level of a learner's interest on learning content as relevant information attracts more attention (Kang, 2014; Rayner, 2009). So students in the experimental group had higher interest on STR-texts as they not only had to comprehend it but able to discuss with their peers as well. This result was also confirmed by questionnaire survey and interview results (please refer to our discussion of questionnaire survey and interview results later in this section).

The results of our investigation suggest that students who attended lectures and discussed their content performed better than those who attended lectures without any discussion. The difference was statistically significant for information recall in particular. Three reasons can explain this finding. First, discussion itself played a valuable role in facilitating learning. Discussion encouraged and stimulated student to take a more active role in their learning (Grzega & Schoner, 2008). During discussion, students articulated a variety of opinions and points of view and acquired knowledge and insight from diverse points of view (Crippena & Earl, 2007). Discussion offered students opportunities to test their ideas and opinions against those of their peers (Aleven & Koedinger, 2002). Discussion also helped students understand and apply what they have learned (Hubscher, 2010). As for using STR-text in discussion, students mentioned that STR-text was useful for their discussion. For example, students used STR-text to refer to important parts during discussion. When students misunderstood some parts of the lecture and then discussed the content, they could find their misconceptions based on STR-text, correct these misconceptions, and improve their understanding of lecturing content. Students said that they have different points of view about lecturing content and when they discussed the lectures they referred to STR-text so that they could understand content deeper and have different perspectives on the content. Our finding is in line with other related research. For example, Noroozi et al. (2013) suggested providing students with the scripts which can be useful for knowledge construction. That is, the scripts can prompt students to paraphrase, criticize, ask meaningful questions, construct counterarguments, and propose argument syntheses during discussions.

Second, preparing to discussion was beneficial for learning. In order to be prepared for discussion, students learning went beyond mere reviewing/listening to learning content (Crippena & Earl, 2007). This is because students not only had to understand learning content but also be able to explain it to other students. As a result, advanced cognitive processes were activated, i.e. students not only reviewed/listened to learning content, but had to think analytically about it, to relate it to what they already know, and to use that knowledge for constructing new knowledge, solving new problems, and addressing new issues too (Hubscher, 2010).

Third, the difference between student performances in two groups was due to the active processing principle. Mayer (2010) argued that learning is an active process which involves selecting, organizing, and integrating learning information based upon prior knowledge for meaningful learning to occur. So students in two groups attended to learning content, mentally organized it into a coherent cognitive representation, and integrated it with prior knowledge. However, students in the experimental group were engaged in active processing of learning material more compared to those in the control group. As a result, performance of students in the experimental group was better compared to their counterparts.

Our ordinary least squares regression results suggested that there were no significant effects of fixation time, fixation count, and EFL proficiency on post-test results. Therefore, we may conclude that there was a direct effect of having discussion remaining on post-test scores, even after controlling for the effects of fixation time, fixation counts, and EFL proficiency.

Our questionnaire results show that most students in the two groups valued STR-text for learning. In addition, most students in the experimental group highly perceived the usefulness of STR-text for discussion. Finally, most students in the two groups expressed their intentions to use STR-text for learning in the future. This finding suggests that most students accepted STR-text for learning. However, few students mentioned that they did not discuss lecturing content with their partners. Noroozi et al. (2012) mentioned that due to some issues, e.g. the lack of social context cues such as physical form, eye contact and group identity, students may lose their interest and willingness to participate in discussion. This issue is very critical and needs to be considered by educators and researchers. In the future, some intervention techniques, such as giving students discussion questions or related issues, need to be introduced to facilitate their discussion. In addition, interaction among students needs to be encouraged.

Although, the questionnaire survey results provide subjective evidence as they are based on student perceptions, we carried out our survey for two main reasons. First, we aimed to measure how our students accept and use technological intervention because several factors contribute to their decision about how and when they will use it (Davis, 1989). For example, perceived usefulness represents the degree to which our students believe that using STR-texts would enhance their learning performance while behavioral intention measures the likelihood of students to use it in the future. So if students have positive perceptions towards STR-texts, they will accept STR-texts and use them for enhancing comprehension of learning content. Second, we aimed to measure student perceptions towards usefulness of STR-texts during learning and discussion in order to be able to triangulate among different data sources (e.g. the results of the questionnaire survey, the post-test results, and interviews) to ground our findings.

Furthermore, we interviewed students to find reasons behind their high level of perceptions towards STR-text. According to students, most of them read STR-text and simultaneously listened to the instructor during lectures. Most students mentioned that this approach helped them understand what the instructor was saying. Some students claimed that their listening skills are not good and therefore they had to rely on STR-text. Other students said that their reading skills are good and it was easier to understand the lectures content by reading STR-text. Some students complained that the instructor's accent was not clear or speech was not fluent so that they couldn't hear some words clearly and had to refer to STR-text for clarifications. Some students argued that it was easy to forget the lecture content after listening to it, but with STR-text students were able to easily recall information which they missed. When we asked students to compare STR-text with lecture slides, most of them mentioned that STR-text contained all lecture information whereas lecture slides only key points. Similar findings about usefulness of STR-text during lectures were reported in other related studies. For example, Hwang et al. (2012) and Kuo et al. (2012) suggested that STR-text can help students attain better understanding of a lecture, do simultaneous note-taking during lectures, and complete homework. Shadiev, Hwang, and Huang (2013) and Wald and Bain (2008) claimed that STR-text is beneficial for learning as it can be used as an additional text confirmation of what is being said, and it aids comprehension in case when listeners are foreign students (Shadiev, Hwang, Chen, & Huang, 2014; Wald & Bain, 2008). The usefulness of STR-text for learning can be explained by a cognitive theory of multimedia learning

(Mayer & Moreno, 2003; Mayer, 2009). Presenting verbal and visual information simultaneously enables students to receive information in two modalities, i.e. verbal (i.e. speech of the instructor) and visual (i.e. STR-text), and to process it through different processing systems. This is why learning content can be better processed and understood.

However, some students mentioned that STR-text was not useful for them. For example, students with lower EFL reading skills mentioned that their reading ability was not good and it took a while to read STR-text. When they read one sentence of STR-text, the instructor already spoke the next one. Thus, using STR-text was not useful and hindered their comprehension. So these students preferred to listen to the instructor instead of reading STR-text. Students, who paid more attention to lecture slides compared to STR-text, mentioned in interviews that they could find key-points or important information regarding the lectures from slides easily. Therefore, lecture slides could help them to understand content better than STR-text. Some students admitted that lecture slides helped their listening to the lectures better than STR-text did. Other students said that lecture slides were useful for later assessment since key points of the slides matched items of the tests. Most students who had preference towards lecture slides were students with better EFL skills. Some students admitted that listening to the instructor was enough for them to understand the lectures content. Most of these students were with higher EFL proficiency. This finding is in line with the expertise reversal principle of cognitive load theory (Kalyuga, 2014). That is, some instructional techniques that are highly effective with one type of learners (e.g. novice) may not be effective and may even have negative consequences, such as increased cognitive load, when used with another type of learners (e.g. more knowledgeable learners). Low ability learners may find information presented in visual and verbal forms as useful for learning due to lack of their ability whereas high ability learners may perceive information presented in two forms as redundant (Clark & Mayer, 2011).

5. Limitations

Some limitations of this study need to be acknowledged and addressed. For example, our sample size was small which limits generalization of the results to a larger population. Another limitation associates with our eye-tracking tool; it could not always detect eye-movement when participants wore glasses. In addition, we investigated how different are effects of discussion on attention, performance, and perceptions of students in general only by comparing post-test results of students in the control and experimental groups. We did not consider their different levels of EFL proficiency or gender. For example, in terms of gender, our participating students majored in applied science and thus, both groups included more males than females. We believe that gender factor could affect research variables of this study. Furthermore, we did not control the language in which students in the experimental group had their discussion. In this study, students used both languages. But it is possible that if they discussed in Chinese or in English only, the entire learning impact would be different due to impact of L1 discussion is different from that in L2. These limitations will be addressed in our future study. We will also focus on other important research variables, such as cognitive load or mental state of students during learning in multimedia environments, in our future study.

6. Conclusions

Based on our results, we may conclude that our experimental treatment was beneficial for visual attention and learning of students in the experimental group. First, fixation time of students in the experimental group during both lectures was significantly higher than that of those in the control group. Second, students in the experimental group had significantly higher fixation count during the first lecture compared to students in the control group. Finally, students in the experimental group outperformed their counterparts on both tests.

Therefore, we suggest providing an opportunity and time for students to discuss learning content during lectures. In this case, discussions will allow students to negotiate meanings, clarify some ideas, and explain important concepts to other students, which, in turn, will facilitate their understanding of lecturing content. We also suggest encouraging students to interact with each other. After some experience to discuss with partners, they may find strength of a discussion and then utilize it more during lectures. In addition, we suggest providing STR-text to compliment other available media (e.g. lecture slides), during lectures to facilitate students' comprehension of lectures content. This approach is especially useful for students with low EFL ability and who lack prior knowledge on the subject. When students experience difficulty comprehending lecturing content, they may refer to STR-text for clarifications, catching up missing details or as a reference for their discussion.

Appendix. Questionnaire survey

I Perceived usefulness of STR-text for learning

- Using STR-text during lecture improves the quality of my learning
- Using STR-text during lecture helps me to accomplish learning tasks more quickly
- Using STR-text during lecture increases my productivity
- Using STR-text during lecture enhances my effectiveness on the learning
- Using STR-text during lecture improves my learning performance
- Overall, I found using STR-text during lecture is useful in my learning

II Behavioral intention to use STR-text

- I intend to continue using STR-text in future
- I plan to use STR-text often

- I will strongly recommend others to use STR-text

III Perceived usefulness of STR-text during discussion

- Using STR-text during discussion improves the quality of my learning
- Using STR-text during discussion helps me to accomplish learning tasks more quickly
- Using STR-text during discussion increases my productivity
- Using STR-text during discussion enhances my effectiveness on the learning
- Using STR-text during discussion improves my learning performance
- Overall, I found using STR-text during discussion is useful in my learning

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147–179.

Chen, I. J., & Chang, C. C. (2009). Cognitive load theory: An Empirical study of anxiety and task performance in language learning. *Electronic Journal of Research in Educational Psychology, 7*(2), 729–746.

Clark, R. C., & Mayer, R. E. (2011). *E-Learning and the science of Instruction: Proven guidelines for consumers and designers of multimedia learning.* San Francisco, CA: John Wiley & Sons.

Crippena, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education, 49*, 809–821.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Grzega, J., & Schoner, M. (2008). The didactic model LdL (Lernen durch Lehren) as a way of preparing students for communication in a knowledge society. *Journal of Education for Teaching, 34*, 167–175.

Huang, Y. M., Liu, C. L., Shadiev, R., Shen, M. H., & Hwang, W. Y. (2015). Investigating an application of speech-to-text recognition: A study on visual attention and learning behaviour. *Journal of Computer Assisted Learning, 31*(6), 529–545.

Huang, Y. M., Shadiev, R., & Hwang, W. Y. (2016). Investigating the effectiveness of speech-to-text recognition applications on learning performance and cognitive load. *Computers & Education, 101*, 15–28.

Hubscher, R. (2010). Assigning students to groups using general and context-specific criteria. *IEEE Transaction on Learning Technologies, 3*, 178–189.

Hwang, W. Y., Shadiev, R., Kuo, T. C. T., & Chen, N. S. (2012). Effects of Speech-to-Text Recognition application on learning performance in synchronous cyber classrooms. *Educational Technology & Society, 15*(1), 367–380.

Jacob, R. J., & Karn, S. K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. H.Radach, & H.Deubel (Eds.). *In the mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Amsterdam: Elsevier Science.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 576–597). New York: Cambridge University Press.

Kang, H. (2014). Understanding online reading through the eyes of first and second language readers: An exploratory study. *Computers & Education, 73*, 1–8.

Kruger, J. L., & Steyn, F. (2014). Subtitles and eye Tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105–120.

Kuo, T. C. T., Shadiev, R., Hwang, W. Y., & Chen, N. S. (2012). Effects of applying STR for group learning activities on learning performance in a synchronous cyber classroom. *Computers & Education, 58*(1), 600–608.

Leitch, D. (2008). *GIFT atlantic liberated learning high school pilot project: A study of the transfer of speech recognition technology from university classrooms to high school classrooms. Phase III Report. Nova scotia*Canada: Saint Mary's University press.

LifeCam HD-6000 (2017). *Microsoft® LifeCam HD-6000.* Available from: http://www.microsoft.com/accessories/en-us/d/lifecam-hd-6000-for-notebooks.

Liu, P. L. (2014). Using eye tracking to understand the responses of learners to vocabulary learning strategy instruction and use. *Computer Assisted Language Learning, (Ahead-of-print),* 1–14.

Mayer, R. E. (2009). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education, 44*(6), 543–549.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*. http://dx.doi.org/10.1002/acp.3050.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

Nisbet, P., Wilson, A., & Aitken, S. (2005). Speech recognition for students with disabilities. *Proceedings of the inclusive and supportive education congress, ISEC 2005 conference*. Delph, UK: Inclusive Technology.

Noroozi, O., Weinberger, A., Biemans, H. J., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCL): A synthesis of 15 years of research. *Educational Research Review, 7*(2), 79–106.

Noroozi, O., Weinberger, A., Biemans, H. J., Mulder, M., & Chizari, M. (2013). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers & Education, 61*, 59–76.

OGAMA (2017). *OGAMA (OpenGazeAndMouseAnalyzer): An open source software designed to analyze eye and mouse movements in slideshow study designs.* Available from: http://www.ogama.net/.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of Human Computer Interaction,* 211–219.

Ranchal, R., Taber-Doughty, T., Guo, Y., Bain, K., Martin, H., Robinson, J., et al. (2013). Using speech recognition for real-time captioning and lecture transcription in the classroom. *IEEE Transactions on Learning Technologies, 6*(4), 299–311.

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

Ryba, K., McIvor, T., Shakir, M., & Paez, D. (2006). Liberated Learning: Analysis of university students' perceptions and experiences with continuous automated speech recognition. *Journal of Instructional Science and Technology, 9*(1), 1–19.

Shadiev, R., & Huang, Y. M. (2016). Facilitating cross-cultural understanding with learning activities supported by speech-to-text recognition and computer-aided translation. *Computers & Education, 98*, 130–141.

Shadiev, R., Huang, Y. M., & Hwang, J. P. (2017). Investigating the effectiveness of speech-to-text recognition application on learning performance, attention, and meditation. *Educational Technology Research and Development*. http://dx.doi.org/10.1007/s11423-017-9516-3.

Shadiev, R., Hwang, W. Y., Chen, N. S., & Huang, Y. M. (2014). Review of speech-to-text recognition technology for enhancing learning. *Educational Technology & Society, 17*(4), 65–84.

Shadiev, R., Hwang, W. Y., Huang, Y. M., & Liu, C. J. (2016). Investigating applications of speech to text recognition for face to face seminar to assist learning of non-native English participants. *Technology, Pedagogy and Education, 25*(1), 119–134.

Shadiev, R., Hwang, W. Y., & Huang, Y. M. (2013). Investigating learning strategies of using texts generated by Speech to Text Recognition technology in traditional classroom. *In Proceedings of the AECT International Conference on the Frontier in e-Learning Research* (pp. 279–286). Taiwan: National Central University & AECT.

Shadiev, R., Wu, T. T., & Huang, Y. M. (2017). Enhancing learning performance, attention, and meditation using a speech-to-text recognition application: Evidence from multiple data sources. *Interactive Learning Environments, 25*(2), 249–261.

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning

environment. *Instructional Science, 40*(2), 297–323.
Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks: Sage Publications.
Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education, 58*(1), 375–385.
Wald, M., & Bain, K. (2008). Universal access to communication and learning: The role of automatic speech recognition. *International Journal Universal Access in the Information Society, 6*(4), 435–447.
Way, T., Kheir, R., & Bevilacqua, L. (2008). Achieving acceptable accuracy in a low-cost, assistive note-taking, speech transcription system. *Proceedings of the IASTED international conference on telehealth and assistive technologies*. ACTA Press.