Tracking learners' visual attention during a multimedia presentation in a real classroom

Fang-Ying Yang*, Chun-Yen Chang, Wan-Ru Chien, Yu-Ta Chien, Yuen-Hsien Tseng

Graduate Institute of Science Education, National Taiwan Normal University, 88 Sec. 4 Ting-Zhou Road, Taipei 116, Taiwan, ROC

ARTICLE INFO

ABSTRACT

The purpose of the study was to investigate university learners' visual attention during a PowerPoint (PPT) presentation on the topic of "*Dinosaurs*" in a real classroom. The presentation, which lasted for about 12–15 min, consisted of 12 slides with various text and graphic formats. An instructor gave the presentation to 21 students whose eye movements were recorded by the eye tracking system. Participants came from various science departments in a national university in Taiwan, of which ten were earth-science majors (ES) and the other 11 were assigned to the non-earth-science group (NES). Eye movement indicators, such as total time spent on the interest zone, fixation count, total fixation duration, percent time spent in zone, etc., were abstracted to indicate their visual attention. One-way ANOVA as well as *t*-test analysis was applied to find the associations between the eye movement data and the students' background as well as different formats of PPT slides. The results showed that the students attended significantly more to the text zones on the PPT slides and the narrations delivered by the instruction. Nevertheless, the average fixation duration, indicating the average information processing time, was longer on the picture zones. In general, the ES students displayed higher visual attention than the NES students to the text zones, but few differences were found for the picture zones. When the students viewed those slides containing scientific hypotheses, the difference in attention distributions between the text and pictures reduced. Further analyses of fixation densities and saccade paths showed that the ES students were better at information decoding and integration.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

According to the Dual-Coding Theory (Paivio, 1986), visual and verbal information are processed in distinct channels. In view of this, it has been widely agreed by educators that teaching or learning materials containing both verbal and visual modes of information should improve learning. Based on various cognitive theories, Mayer and colleagues proposed the theory of multimedia learning (Mayer, 2001, 2005; Mayer & Sims, 1994) which guides multimedia instructional design. The number of studies regarding multimedia learning has grown significantly in recent years. However, since most studies dealing with multimedia learning are conducted in experimental settings, there still exists uncertainty about how learners process multimedia information in real classrooms. To probe in-depth into how students learn concepts in the science classroom with multimedia materials, we conducted a study that examined students' visual attention in terms of their eye-movement patterns as they were given a multimedia presentation using the Microsoft PowerPoint software in the classroom. The following two issues were explored in the study: (1) In a real classroom, how would university students distribute their visual attention to a multimedia presentation with different text–picture formats? (2) How do learners with different backgrounds differ in their processing of multimedia material?

---

2. Literature review

2.1. Theory of multimedia learning

The theory of multimedia learning proposed by Mayer and colleagues (Mayer, 2001; Mayer & Sims, 1994) was constructed based on cognitive theories such as the dual code theory (Paivio, 1986) and the cognition-load theory (Plass, Moreno, & Brunken, 2010; Sweller, 1988). The dual code theory argues that human cognition has two distinct but interconnected systems that process different modes of information (i.e., visual and non-visual modes) in different channels. Studies supporting the dual system have shown that when learners were instructed to form images as they read instructional texts, or were cued by pictures, their recall of the facts or concepts presented in the texts was improved (e.g., Paivio, 2006; Paivio & Lambert, 1981). It was also found that combinations of text, imagery and pictures in instruction facilitated conceptual understanding (e.g., Purnell & Solman, 1991; Sadoski & Willson, 2006). Cognitive load theory (Paas, Renkl, & Sweller, 2003) is an instructional theory that concerns the interaction between instructional information and memory structures. It states that the way information is presented, the learning activities, and the element interactivity of the information together impose a cognitive load on learners. This cognitive load could be intrinsic, extraneous (ineffective), or germane (effective), depending on the nature of the learning material and the instructional design. The total load cannot exceed the capacity of working memory if optimum learning is expected.

Incorporated with the above-mentioned cognitive theories, Mayer and colleagues (Mayer, 2001; Mayer & Sims, 1994) proposed the theory of multimedia learning, which emphasizes the role of experience and ability in learning from various nonverbal representations including pictures, animations and narrations. In addition, taking the idea from the generative theory (Wittrock, 1989), Mayer further (1997, 2005) pointed out that meaningful learning in multimedia environments occurs when learners select relevant information, organize the information to form coherent mental representations, and integrate the new and existing representations. Some instructional principles have been introduced based on the theory, including multiple representations, contiguity, split-attention, coherence principles, modality, individual differences, and so forth (Mayer, 2008). These principles to date have become the major guidelines for designing multimedia instruction.

2.2. Effects of multimedia learning

Research in multimedia leaning has been growing during the past two decades due to the rapid development of educational technologies. Although numerous studies have shown positive effects of multimedia instruction designed based on the theory of multimedia learning (e.g., Mayer, 2005, 2008; Reisslein, Seeling, & Reissien, 2005), many studies have also demonstrated exceptions to or concerns about factors affecting the results of multimedia instruction. For example, Dillon and Jobus (2005) reviewed the literature on hypermedia learning since 1998, and found that the results were mixed. Recently, a study conducted by Change, Lei, and Tseng (2011) found modality but no redundancy effects in the learning of foreign language when the text and audio modes of information were presented together. Leslie, Low, Jin, and Sweller (2012) found that between audio–visual and audio only presentations, the latter was beneficial for older students with prior knowledge of the topic to be learned, but younger students with no prior knowledge learned better from the audio–visual form. Some other studies showed that multimedia materials were effective when incorporated with cuing instructional strategies, and/or took into consideration the learning pace and learner characteristics such as prior knowledge, cognitive styles and personal theories (e.g., Greene, Costa, Robertson, Pan, & Deekens, 2010; Hoffler & Schwartz, 2011; Liu, Andre, & Greenbowe, 2008; Pastore, 2012; Yang & Chang, 2009).

The inconclusive research findings in multimedia learning have aroused a great deal of discussion. Most scholars agree that other than the instructional design, the effectiveness of multimedia learning is also mediated by factors such as the context, the goal of learning and individual differences (Mayer, 2008; Dillon & Jobus, 2005). How these factors contribute to learning in multimedia environments needs thorough examination. Some researchers recognize that many studies of multimedia learning were actually conducted in laboratories or experimental settings, and consequently, whether the positive results can be transferred to the classroom environment remains uncertain (Rieber, 2005). In our view, to gain an in-depth understanding of the effect of multimedia learning, and to provide insights regarding the variety of research findings discussed above, multimedia research must include studies examining how multimedia information is processed.

2.3. Analyzing the process of multimedia learning using eye tracking technology

Traditionally, the interview method using the think-aloud protocol has been the most important and frequently used technique for probing the process of learning (Mintzes, Wandersee, & Novak, 2000). However, such a method often suffers from the uncertainty existing in introspective thoughts. For that reason, educational researchers have been seeking different research methods in the hope of presenting the process of learning from different perspectives. In psychology, the eye tracking system has been used for decades to study cognitive processes. In particular, this technique has been intensively employed in reading and information processing research (Rayner, 1998). The eye tracking method is known for its capacity to record the online cognitive processes and reveal the basic mechanisms of information decoding and integration. Based on the eye-mind assumption that eye fixation locations reflect attention distributions (Just & Carpenter, 1980), the eye tracking method can reveal the temporal change of visual attention that may further inform how learners approach and process information during learning. However, due to little communication between researchers in education and psychology, such a research tool did not receive enough attention from education researchers until recent years.

Using the eye tracking technique as a tool to reveal the process of learning is a new and developing research approach in the domain of education. Related literature in the area of multimedia learning is beginning to grow. There have been some studies using the eye tracking method to explore learners' attention distributions with respect to different components of learning material, such as text, graphics, illustrations and so forth, to reveal how learners spend their cognitive resources on the multimedia information (e.g., Hyönä, 2010; Liu & Chuang, 2011). An interesting finding from these studies is that text seemingly attracts most of learners' attention. Other studies have employed the eye tracking method to examine the effects of multimedia as indicated by the theory of multimedia, such as redundancy, modality and contiguity (e.g., Rummer, Schweppe, Furstenberg, Seufert, & Brunken, 2010; Schmidt-Weigand, Kohnert, & Glowalla, 2010; Liu,

Lai, & Chuang, 2011). These eye movement analyses reveal mixed results of the effects during multimedia learning. In short, the eye-tracking technology allows researchers to picture how learning material is attended to, and to probe in-depth the interactions among processing behaviors, instructional design and learning outcomes.

2.4. The role of prior knowledge in the process of multimedia learning

By examining the associations between learners' prior knowledge and learning outcomes, cognitive and educational studies have shown that prior knowledge not only affects the recall and comprehension of domain-related information, but also mediates conceptual change (e.g., Carey & Spelke, 1994; Schneider & Korkel, 1989; Walker, 1987). As mentioned previously, many researchers who have examined the effect of prior knowledge in the context of multimedia learning have demonstrated the interaction between prior knowledge and learning outcomes (e.g., Greene et al., 2010; Liu et al., 2008; Mayer, 2008), but few studies have focused on the learning process. Using the eye tracking method, some recent studies have shown that learners with different levels of prior knowledge display different processing approaches (e.g., Jarodzka, Scheiter, Gerjets, & Van Got, 2010; She & Chen, 2009). Nevertheless, how these processing behaviors are similar or differ across different subject matters has yet to be fully understood. Thus, another research task in this study was to explore the issue regarding the role of prior knowledge.

In multimedia environments, pictures or graphics are the main source of information for learning. Although how viewers read scenes has attracted great attention from researchers in psychology, there are still debates over what factors influence scene reading (Foulsham, Kingstone, & Underwood, 2008; Rayner, 2009). Earlier studies suggest that saliency in a picture (such as contrast, colors, intensity, spatial frequency, etc.) guides where viewers look (e.g., Koch & Ullman, 1985), but later investigations using the eye tracking method argue that there are strong high-level cognitive influences (e.g., Humphrey & Underwood, 2009, 2010; Neider & Zelinsky, 2006; Underwood, Foulsham, van Loon, & Underwood, 2005). Prior knowledge, in particular domain knowledge, is identified as being a significant cognitive factor mediating the visual attention during scene viewing (Humphrey & Underwood, 2009; Tatler, Baddeley, & Gilchrist, 2005). However, since the studies of scene perception have employed mostly natural photos or real-life pictures, how learners read the knowledge-based pictures that are frequently included in science texts needs further clarification. Thus, in this study, an attempt was made to explore this issue.

2.5. Issues and research questions

Two conclusions can be drawn from the above literature review. First, although research in multimedia learning has been growing significantly in recent years, there is still a lack of studies conducted in authentic classrooms. Second, as many of the existing studies show inconsistent findings, different research approaches are needed to provide insights to explain this discrepancy. Thus, in this study, we have made an attempt to examine learners' visual attention distributions using the eye tracking method as they learned science concepts through a PowerPoint (PPT) multimedia presentation in a real classroom. Two research questions were examined:

1. When viewing a PowerPoint (PPT) multimedia presentation consisting of various text–picture formats in a real classroom where an instructor narrated the content, how did the university students allocate their visual attention?
2. Was there an effect of prior knowledge on the visual attention distributions?

3. Method

3.1. Subjects

Twenty-one university students in various science departments, including earth sciences, physics, chemistry and biology, in a national university in Taiwan were voluntarily involved in the study. These students were either in the sophomore or junior years. Among the participants, ten who were majoring in earth sciences were labeled as the ES students, while the other 11 were assigned to the "non-earth-science" (NES) group. It should be noted that the department of Earth Sciences in the participating university offers programs in various disciplines including geology, geophysics, oceanography, meteorology and astronomy. Students in the freshman and sophomore years are required to take the introductory courses in each discipline. Accordingly, the ES students, who had already taken courses related to the theory of plate tectonics and the history of the earth, were identified as learners with relevant prior knowledge. On the other hand, the NES subjects did not take any formal courses related to the earth science.

3.2. Material

A 12–15 min PPT presentation on the topic of "Truth about Dinosaurs" was prepared for the study. The PPT lesson consisted of 12 slides showing various text–picture formats. Table 1 shows the design of each slide.

As Table 1 shows, the first slide gives the outline of the lesson. Three slides (slides# 2, 3 and 6) contain photos only, while 5 (slides# 4, 5, 7–9) have both text and picture components. Two text–picture slides (slides# 8 and 9) have highlights on the photos (one static highlight and the other animated). The last three slides describe three popular hypotheses about dinosaur extinction. Rather than photos, the three conceptual slides contain different types of conceptual graphics. The content and the design of the PPT presentation were constructed by a content expert who was a professor of the relevant area and a science education researcher. We ensured that the photos used in the PPT material did not appear in any textbooks that the participating ES students might have read prior to the study.

3.3. Instruction

The PPT lesson was given to the participants, one at a time, in order to record their eye movements. When giving the lecture, the instructor started with the blank outline slide with only a title on it, and then asked the participating student to think about evidence he/she

Table 1
The design of the PPT presentation.

PPT slide	Text–picture design	Initiation
1. *How can you know there were dinosaurs?*	Outline slide indicating sub-topics of the lesson	Teacher
2. Bone	A genuine photo	Student
3. Foot print	A genuine photo	Student
4. Nest	Text and a genuine photo	Student
5. Fossil eggs	Text and a genuine photo	Student
6. *How can you know there were dinosaurs?*	Outline slide indicating sub-topics of the lesson	Teacher
7. Coprolites	A genuine photo	Teacher
8. Tooth marks	Text and a genuine photo	Teacher
9. Teeth	Text and a genuine photo with an animated highlight	Teacher
10. Skin marks	Text and a genuine photo with a static highlight	Teacher
11. Cause of dinosaur extinction (Asteroid impact)	Text and a conceptual graphic (A planet model)	Teacher
12. Cause of dinosaur extinction (Plate tectonics)	Text and a conceptual graphic (A structural model of plate tectonics)	Teacher
13. Cause of dinosaur extinction (Climate change)	Text and a numerical table	Teacher

knew of for the existence of dinosaurs. In doing so, the student was encouraged to initiate some relevant content. Four types of evidence that students would be most familiar with were expected to be identified at this stage. The instructor gave hints if the subject had difficulty pinpointing the evidence. Subsequently, a brief summary of the evidence was shown on the outline slide. The instructor then went over the student-initiated content presented on different slides. Afterward, the instructor went back to the outline slide and talked about other evidence that the students might not be so familiar with. Once again, another summary was shown after the unfamiliar evidence was introduced. The instructor then went over these related slides. Consequently, the outline slide appeared twice during the lecture, separating the student- and teacher-initiated ideas, as indicated in Table 1. Afterward, the instructor went on to explain the three hypotheses about dinosaur extinction displayed in the last three slides. Noticeably, the text zones of the text–picture slides contain information identical to the narrations delivered by the instructor.

3.4. Apparatus

This study employed the *faceLAB 4.5* eye tracking system developed by Seeing Machines Company, Australia, which is a remote, non-intrusive and fully automated eye and head tracking system. The system consists of one laptop computer and two miniature cameras that allow the detection of depth. It is able to provide data such as eye movements, pupil size, blink rate, and head movements. According to Larsson (2002), the faceLAB system employs a mix of different algorithms to track gaze direction. It first tracks the position of the subject's head using eye-brows, nose and lips, and then uses the data to track the subject's gaze. The identification of fixation is determined by the analytic software, GazeTracker, which is introduced later. The average sampling rate of faceLAB is 60 Hz, that is, 60 eye-movement samples are captured in 1 s. The typical static accuracy of gaze direction measurement is 0.5–1° of rotational error. In addition, it has been reported that people can generally demonstrate a precision better than 0.5° of visual angle. This system has been widely used in human factor, human performance and simulator-based studies (e.g., Fletcher, Loy, Barnes, & Zelinsky, 2005; Young, Mitsopoulos-Rubens, Rudin-Brown, & Lenne, 2012). Recently, the apparatus has also been applied in some multimedia learning studies (e.g., Liu et al., 2011; Tsai, Hou, Lai, Liu, & Yang, 2011).

3.5. Procedure

The study was set up in a real classroom. Since there was only one set of eye-tracking equipment available for use, the lecture was given by the same instructor 21 times so that the eye-movements of each participant could be collected. Before the study, the instructor underwent a training session to make sure that he would give the PPT presentation at a proper and consistent pace, and deliver the same quality of content about the topic each time. The multimedia presentation was projected on a 180 × 150 cm screen on the wall, and the screen was 4.6 m from where the subject was seated. The eye tracker was then placed on a table in front of the subject about 70 cm away. Each participating subject went through a calibration process for the eye tracker to capture the correct positions of the subject's eye movements. All the participants passed the calibrations with an accepted angular error of less than 1.0° measured by the eye tracker faceLAB 4.5. The whole presentation took about 12–15 min to complete. During the lecture, a real-time monitoring and recording system was utilized for the researcher to constantly scrutinize the subjects' eye movements. This system overlaid both the subject's eye movements and the PPT material on a computer screen. It also simultaneously recorded the instructor's narrations. According to the post-hoc analysis of the overlaid records, all the subjects' eye movements followed the instructor's narrations, and most of their fixations fell in the areas of interest. At the end, students took a free-recall memory test. Two raters worked together to analyze the number of concept terms and related propositional ideas recalled by these participants. The analysis showed that there was no difference in the recall of the concept terms, but the ES students were able to state significantly more propositional ideas discussed in the lecture than the NES students did (Mean score = 12.73, SD = 3.26 for the ES group, while Mean score = 9.00, SD = 4.45 for the NES group, $F = 4.87, p < 0.05$).

3.6. Data analysis

The eye movement patterns were analyzed by the *GazeTracker* 8 software which classifies fixations based on the dispersion-threshold identification algorithm that utilizes the fact that fixation points tend to cluster closely together (Salvucci & Goldberg, 2000). The maximum gap interval between two subsequent gaze points that can be included in a fixation is set to 0.035 s. According to Rayner's review (1998), fixation durations may range from 100 ms to 500 ms, with an average of about 250 ms. Tsai, Yen, and Wang (2005) also found similar average fixation durations among Chinese readers. In addition, during scene perception, fixation durations tend to be longer than those in text reading, with the average fixation duration being closer to 300 ms (Rayner, 2009). Although an eye fixation can pick up information of the

Fig. 1. An example of eye movement patterns of one subject. The square areas indicate look-zones (i.e., areas of interest).

simple visual stimulus in less than 100 ms, prior studies have shown that about 150–200 ms is needed for our brain to interpret and utilize the fixated information (Salthouse & Ellis, 1980; Sereno, Rayner, & Posner, 1998; Vaughan & Graefe, 1977). Considering that the main inquiry of this study is related to the reading of conceptual passages and graphics, we focused on analyzing fixations with a duration greater than 150 ms. To study the attention distribution, the various fixation measures, indicating attention distributions and time needed for information processing, and saccade paths, suggesting sequence of information processing and integration of information, were the main foci of the analysis.

For the purpose of examining the subject's attention distributions on the different components of the PPT slides, each slide was divided into several 'look-zones' (as indicated by the square areas shown in Fig. 1) consisting of title, texts, picture or graphics with or without animation.

To summarize the eye movement patterns on each PPT slide, 3 eye-movement measures were used, including total fixation duration (TFD), number of fixations (NF), and average fixation duration (AFD). Besides, the total time shown (TTS) and the total time tracked (TTT) were also abstracted for calculating the percentage of various attention distributions. Meanwhile, to analyze the attention distributions on different medium components (look-zones) on a slide, 6 eye movement measures were used, including percentage of time spent in zone (PTSZ), fixation count (FC), percentage of total fixations (PTF), total fixation duration (TFD), percentage of time fixated related to total fixation duration (PTFRTFD), and average fixation duration (AFD). These eye movement measures represent cognitive activities related to reading, comprehension and movement of attention. In brief, the fixation measures such as NF, FC and TFD, indicate the period of time needed to acquire new information (Rayner, 2009). Meanwhile, the average fixation duration, AFD, while reveals the time for information processing, could be influenced by the nature of the task given to the participants (Rayner, 2009). On the other hand, the percentage measures including PVT, PTSZ, PTF and PTFRTFD were employed in the study to show attention distributions in terms of reading time and fixation durations for different target areas of interest. In addition, the times of saccade paths indicating the back-and-forth scanning between different zones were recorded. These times can reveal the processes of integration between different modes of information (Holsanova, Holmberg, & Holmqvist, 2009). Table 2 provides the definitions of these eye movement and other measures.

The eye-movement data were exported to Excel, and SPSS was then applied for further statistical analyses. The descriptive statistics, *t*-tests and one-way ANOVA along with the homogeneity tests were performed to find differences in eye movements between the slides and between the different background groups (ES vs. NES). The *t*-tests showed the within-subject difference concerning variation between different modes of presentation, while one-way ANOVA was applied to examine between-subject differences regarding the students' different academic backgrounds. Moreover, to illustrate the difference of visual attention to the conceptual models among learners with different academic backgrounds, the fixation densities of the two academic groups on the last three slides were collected and compared.

Table 2
Definitions for the eye-movement measures and other study measures.

Eye-movement measure	Definition
1. Total fixation duration (TFD)	Sum of durations of all fixation points on a slide
2. Number of fixations (NF)	Sum of number of all fixation points on a slide
3. Average fixation duration (AFD)	Average duration of a fixation point
4. Percentage of viewing time (PVT)	Total fixation duration divided by the total time shown
5. Percentage of time spent in zone (PTSZ)	Total time in a look-zone, such as a text or picture zone, divided by total time tracked
6. Fixation count (FC)	Number of fixation points in a zone
7. Percentage of total fixation (PTF)	Fixation count divided by number of fixations
8. Percentage of time fixated related to total fixation duration (PTFRTFD)	Total fixation duration in a zone divided by total fixation duration of the whole slide
9. Sum of saccade paths (SSP)	Sum of all back-and-forth scanning (saccade) between look-zones
10. Frequency of saccade path (FSP)	Times of saccades divided by total time tracked
Other measure	**Definition**
1. Total time shown (TTS)	Total time displayed by a slide, including time not recorded by the eye tracker
2. Total time tracked (TTT)	Total time in a slide recorded by the eye tracker, including fixation and saccade durations

4. Results

To avoid long explanations of each eye movement measure, the findings of the study are organized in response to the research questions, focusing on how the various slides as well as the text/picture zones of the multimedia presentation are viewed and inspected. A discussion of the role of prior knowledge then follows.

4.1. Eye movement patterns with respect to different text–picture formats and look-zones

Table 3 displays a brief summary of the students' general eye-movement patterns for each PPT slide. Paired-t tests were applied to test the differences between slides of different text–picture formats. As Table 3 shows, the percentages of viewing time (i.e., the total fixation duration divided by the total time tracked) on the PPT slides ranged from 29.7% to 59.2%, meaning that much of the students' attention was allocated outside the PPT presentation. It is believed that the instructor was the main target of attention when the students were not attending to the PPT presentation. The outline slide received the least percentage of viewing time at its first appearance. On average, the slides containing both text and photos seemed to receive higher visual attention than those showing photos alone (average PVTs = 43.8% and 35.1% respectively, $t = 4.1$, $p < 0.01$). Nevertheless, the average fixation durations related to the time needed for processing information, were longer for the slides with photos alone (average AFD = 261 ms for the simple text–photo slides and 280 for the photo-alone slides, $t = 3.58$, $p < 0.01$), especially those slides related to ideas initiated by the students (e.g., slides about bones and footprints). Among the three text–photo slides (Nest, Egg and Tooth Marks), the average fixation durations and the percentages of viewing time were slightly higher for the one raised by the instructor, but no statistical significance was found. In addition, among all the text–picture slides except the last three conceptual ones, higher average fixation durations were found for the two with highlights (AFD = 273 ms for highlighted slides, and 261 ms for the others, $t = 2.70$, $p < 0.05$). Further, between the text–photo-highlighted slides (i.e., teeth vs. skin mark), the latter received a higher percentage of viewing time ($t = 3.69$, $p < 0.01$), but the average fixation durations showed no difference.

It has been mentioned that the last three slides discussed the hypotheses about dinosaur extinction. Three different types of pictures were employed. It was found that the participants spent the highest overall visual processing time, indicated by the average fixation duration, on the "Asteroid impact" slide. Regarding the percentage of viewing time, no statistical difference was found between the "Asteroid impact" and "Plate tectonics" slides. However, the students showed the lowest percentage of viewing time when reading the slide describing climate change as the cause of extinction (compared with "Asteroid impact", $t = 3.60$, $p < 0.01$; with "Plate tectonics, " $t = 2.5$, $p < 0.05$).

4.2. Visual attention distributions for different look-zones

As previously mentioned, three areas of interest including title, text and picture were denoted as look-zones for analysis. Table 4 lists a summary and the result of the t-test analyses for attention distributions, indicated by PTSZ, FC, PTFRTFD and AFD, on the different look-zones of individual PPT slides. It should be noted that, since titles received rather low reading time, they were not included in the t-test analysis. The result of the analysis showed that, first, students rarely looked at the title zones. Second, when text and pictures appeared together on a slide, higher visual attention was allocated to the text zones, but the average fixation durations on the pictures were generally higher. Third, among all the text–picture slides, the conceptual graphics seemed to receive higher percentages of visual attention compared to that given to actual photos, except for the photo of "Teeth," which included an animated highlight. Fourth, between slides with high-lighted pictures, the attention distribution was higher for the animated picture; in contrast, the text zone of the slide with the static highlight received higher attention. Among the last three conceptual slides, a significant difference between the text and picture zones was found only on the plate tectonics slide. Meanwhile, the average fixation duration on the "Plate tectonics" graphic was also the highest.

4.3. Attention distributions and academic background

To find any differences of eye movement patterns between the different academic backgrounds, one-way ANOVA was applied. Homogeneity tests were first conducted to examine the equal variance requirement. Table 5 displays the summary of overall significant effects. It should be noted that since there were only about 10 subjects in each academic group, the statistical effects could be small. Therefore, rather than merely reporting the findings at the 95% confidence level, findings significant at the 90% level are also noted. According to Table 5, the

Table 3
Summary of attention distribution on each PPT slide.

PPT page (Student or Teacher initiated)	Text–picture format	TTS/TTT (s)	TFD (s) (PVT %)	NF	AFD (ms)
How can you know there were dinosaurs? (S)	Outline page	20.0/8.7	6.0 (29.7)	22.8	255
How can you know there were dinosaurs? (T)	The outline (Second appearance)	8.7/7.2	4.9 (59.2)	35.7	269
Bone (S)	Photo alone	54.4/24.6	16.8 (32.4)	58.9	278
Foot print (S)	Photo alone	62.6/27.7	20.4 (35.8)	68.5	292
Coprolites (T)	Photo alone	45.8/23.4	16.8 (37.1)	60.7	270
Nest (S)	Text–picture	37.4/25.0	16.4 (42.5)	62.9	257
Fossil eggs (S)	Text–picture	66.3/38.7	24.1 (37.2)	92.9	259
Tooth marks (T)	Text–picture	41.7/30.8	20.7 (49.0)	77.0	266
Teeth (T)	Text–picture-highlight (animated)	38.5/22.3	14.5 (40.8)	53.0	271
Skin marks (T)	Text–picture-highlight (static)	50.2/37.4	25.8 (50.9)	93.1	274
Cause of extinction (Asteroid impact) (T)	Text–conceptual graphic (planet model)	18.8/12.4	8.5 (49.9)	30.1	287
Cause of extinction (Plate tectonics) (T)	Text–conceptual graphic (Cartoon drawing)	27.2/18.3	12.7 (46.4)	45.5	274
Cause of extinction (Climate change) (T)	Text–numerical diagram	25.0/13.4	9.1 (38.5)	33.5	270

Note: (1) TTS: Total time shown; (2) TTT: Total time tracked; (3) Percentage of viewing time (PVT) = Total fixation duration (TFD)/Total time shown (TTS); (4) NF: Number of fixations; (5) AFD: Average fixation duration.

Table 4
Distributions of visual attention on different look-zones and the significance of the t-test analyses for the distributions between text and picture zones.

PPT slide (S: Initiated by students)	PTSZ (%)			FC			PTFRTFD (%)			AFD (ms)		
	Title	Text	Picture	Title	Text	Picture	Title	Text	Picture	Title	Text	Picture
How can you know there were dinosaurs? – First appearance	4.2	13.3	18.8	2.2	7.00	10.2(*)	8.7	28.7	47.5*	232	235	264(*)
How can you know there were dinosaurs? – Second appearance	0.4	40.9**	3.3	0.1	12.0**	1.4	0.5	51.6**	3.9	258	359**	245
Bone (S)	0.0	N/A	38.2	0.0	N/A	53	0.0	N/A	84.3	N/A	N/A	275
Foot print (S)	0.9	N/A	34.8	1.6	N/A	58.0	2.3	N/A	61.7	284	N/A	284
Coprolites	3.4	N/A	39.9	3.8	N/A	45.0	6.3	N/A	71.2	235	N/A	274
Nest (S)	0.7	42.2**	15.2	0.8	39.2**	14.4	1.0	57.5**	25.3	209	256	277(*)
Fossil eggs (S)	0.7	28.3*	20.8	1.3	45.4*	33.7	1.5	48.4(*)	36.8	250	254	260
Tooth marks	1.8	42.4**	17.5	1.9	46.5**	17.0	2.9	55.4**	26.7	244	254	294**
Teeth with animated highlight	1.1	14.2	34.0	0.8	11.6	30.5	1.8	22.0	59.1**	254	244	283**
Skin marks with static highlight	1.1	43.7**	20.3	1.4	55.5**	26.8	1.8	57.7**	29.5	232	268	287
Cause of extinction (Celestial cause)	0.8	24.9	36.5	0.2	13.1	12.8	0.6	37.9	47.6	172	269	286
Cause of extinction (Plate tectonics)	0.7	37.9**	21.4	0.6	26.3**	14.0	1.1	57.4**	31.8	228	265	300*
Cause of extinction (Climate change)	1.2	20.2	25.6	0.6	12.4	15.2	1.6	40.9	41.0	210	279	275

Note: (1) PTST: Percent time spent in zone; (2) FC: Fixation count; (3) PTFRTFD: Percentage of time fixated related to total fixation duration; (4) AFD: Average fixation duration; (5) (*) $p < 0.1$; *$p < 0.05$; **$p < 0.01$.

associations between eye movement measures and academic background were evident for most slides, for which the ES students seemed to pay higher visual attention to the text zones. However, for the picture zones, either no background effect was found or higher visual attention was actually found among the NES students. As far as the three slides discussing the possible cause of dinosaur extinction are concerned, the background effect was also more significant for the text zones. It should be noted that, while no background effect was found for the pictures illustrating the asteroid impact and climate change hypotheses, the NES students showed higher visual attention to the conceptual graphic of the plate tectonics slide. Nevertheless, the effect was only significant at the 90% level of confidence.

4.4. Fixation density analysis of slides describing "dinosaur extinction"

In the PPT presentation, most of the slides described facts about dinosaurs, and actual photos were used to demonstrate these facts. Unlike these slides, the last three explaining scientific hypotheses about dinosaur extinction were illustrated with three different conceptual

Table 5
Effects of prior knowledge on slides and different look-zones.

PPT page	Design	Look-zone	Significant differences in eye movement measures	Main effects
How can you know there were dinosaurs? (First appearance)	Outline slide (text and photo)	Summary		Not found
		Text	PTF, PTFRTFD, AFD	ES > NES
		Photo	PTFRTFD*	ES < NES
How can you know there were dinosaurs? (Second appearance)	Outline slide (text and photo)	Summary	TTT, NF	ES > NES
		Text	FC	ES > NES
		Photo		Not found
Bone (S)	Photo alone	Summary	TFD*, NF*	ES > NES
		Photo	PTSZ*	ES > NES
Foot print (S)	Photo alone	Summary		Not found
		Photo	TTZ, FC, PTF, TFD, PTFRTFD	ES > NES
Corllopus	Photo alone	Summary		Not found
		Photo	PTF*	ES > NES
Nest (S)	Text and photo	Summary	TFD, NF	ES > NES
		Text	FC, TFD, PTFRTFD	ES > NES
		Photo		Not found
Fossil eggs (S)	Text and photo	Summary		Not found
		Text	FC*, PTF*, TFD, PTFRTFD	ES > NES
		Photo		Not found
Tooth marks	Text and photo	Summary	TTT*, TFD, NF	ES > NES
		Text	PTSZ*, FC, PTF, TFD, PTFRTFD, AFD	ES > NES
		Photo		Not found
Teeth	Text and photo with an animated highlight	Summary	TTT*, NF*	ES < NES
		Text	PTSZ, PTF*, TFD*, PTFRTFD*	ES > NES
		Photo		Not found
Skin marks	Text and photo with a static highlight	Summary		Not found
		Text	PTF*, PTFRTFD, AFD	ES > NES
		Photo		Not found
Cause of extinction (Asteroid impact)	Text and a conceptual graphic (planet model)	Summary	TTT, TFD, NF*, AFD	ES > NES
		Text	PTSZ, PTF*, AFD	ES > NES
		Graphic		Not found
Cause of extinction (Plate tectonics)	Text and a conceptual graphic (cartoon picture)	Summary		Not found
		Text	PTSZ, PTF*, TFD*, PTFRTFD	ES > NES
		Graphic	PTFRTF*	ES < NES
Cause of extinction (Climate change)	Text and a numerical diagram	Summary		Not found
		Text	FC*, PTF*, TFD, PTFRTFD, AFD*	ES > NES
		Numerical table		Not found

Note: The asterisk sign (*) indicates 90% level of confidence. For the rest, the 95% confidence level was achieved.
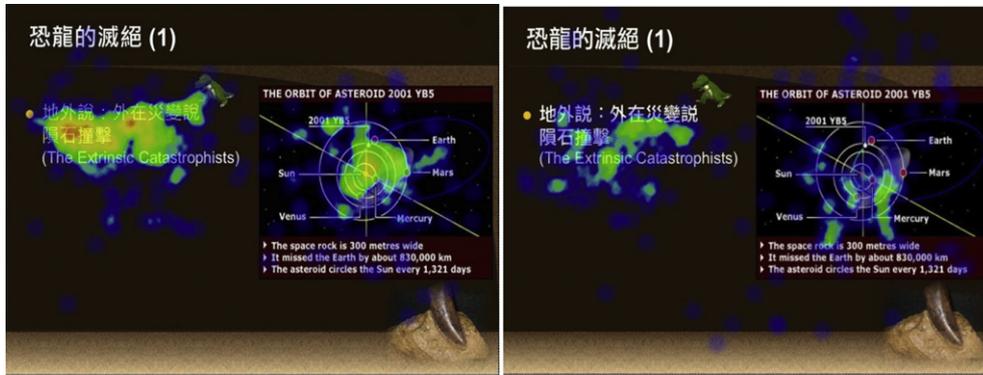
Fig. 2. Comparison of eye movement patterns on the slide of "*Asteroid impact*" between SE (on the left) and NSE (on the right) students. As the graphics show, the fixation densities of the SE students were higher on keywords in the text zone and on the orbit center as well as on the labels in the graphic zone.

graphics. These different types of graphics, including two different conceptual models and a numerical graphic, are frequently seen in science reports or textbooks. Therefore, an in-depth analysis of these slides shall shed some light on how students approach similar scientific illustrations. It has been shown that when viewing the three slides, the ES students paid more attention to the text zones than did the NES students. However, no significant difference was found in the graphic zones, except for the plate tectonics slide. This finding seems to suggest that prior knowledge was associated more with the text reading rather than with the graphic viewing.

To further probe into how students of different backgrounds inspected the different hypotheses, we accumulated all the fixation points of the subjects of the same background on each hypothesis slide. By this method, exactly which location on each slide received the most visual attention overall can be displayed visually. Figs. 2–4 show the result.

Several findings can be drawn from these figures. First, the ES students attended more to the keyword areas in the text zones. For instance, the fixation densities for the ES students were higher for the key terms such as "asteroid impact," "plate/volcanic activities," and the "living environment" of the individual hypotheses. Second, the fixations of the ES students were more concentrated on some particular graphic areas. For example, on the "Asteroid impact" slide, fixation density was higher for the ES subjects at the center of the orbiting planets and the labels of the photo. When viewing the graphics about plate tectonics, the ES students tended to look more at the plate boundaries and the mountain areas where volcanic activities were evident. Nevertheless, the eye movements were more arbitrary on the "climate change" slide. Even so, there was seemingly a trend that the ES students appeared to notice the low peaks of the numerical line more, indicating fluctuation of past temperatures.

4.5. Analysis of saccade paths

Students' back-and-forth scanning (saccade paths) between different look-zones was calculated and the results are displayed in Table 6. The sum of saccade paths (SSP) indicates the total number of times of back-and-forth scanning between different look-zones, while integration between text and picture (ITP) is the sum of the saccade paths between the text and picture zones. Comparisons of the saccade paths between different background students were performed by ANOVA. However, it should be noted that the students did not spend the same amount of time on each slide. Consequently, the longer the viewing time of a slide, the higher the chance of performing saccade scanning. Hence, to reduce the possibility of overestimating students' back-and-forth scanning, we divided the SSP on each slide by the total time tracked (TTT) of the slide. The re-calculated data are listed in parentheses in Table 6. The ANOVA was then conducted. The manipulation would give us some idea about how frequently the saccade scanning was performed by different participants in the same given period of time.

The saccade path numbers show that inter-zone scanning was evident during the PPT learning, and the mean comparisons including ANOVA seem to suggest a trend that the ES students performed higher back-and-forth scans. A similar result was obtained from the re-calculated data indicating the frequency of scanning per second. As far as the three conceptual slides are concerned, significant differences were found in the "asteroid impact" and "plate tectonics" slides, but there were no differences in the "climate change" slide.
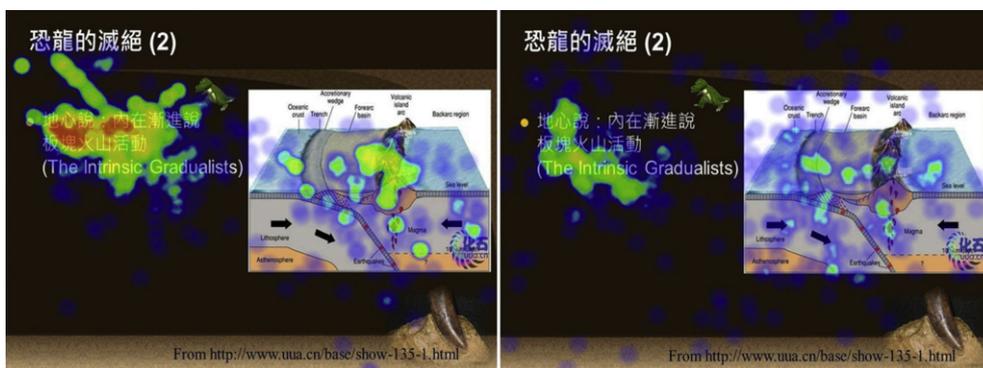

Fig. 3. Comparison of eye movement patterns on the slide of "*Plate tectonics*" between SE (on the left) and NSE (on the right) students. As the graphics show, the fixation densities of the SE students were higher on keywords in the text zone and on the plate boundary as well as on the mountain with volcanic activity in the graphic zone.
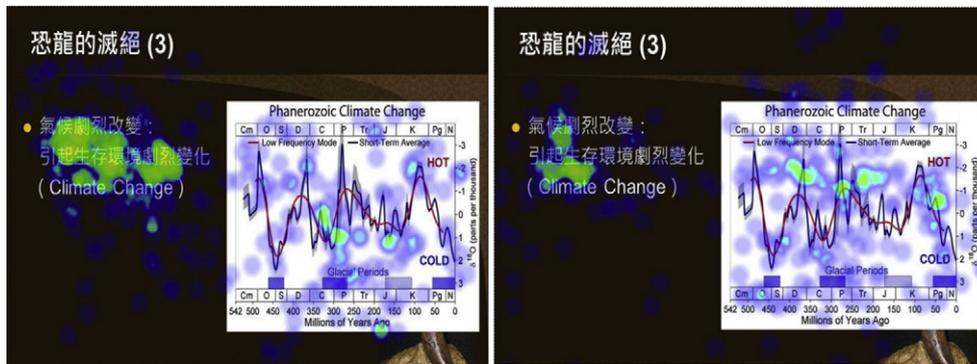
Fig. 4. Comparison of eye movement patterns on the slide of "*Climate change*" between SE (on the left) and NSE (on the right) students. As the graphics show, the eye movement patterns of both the SE and NSE students are more arbitrary. Nevertheless, the SE students seem to have noticed more of the low peaks of the data line.

5. Summary and discussion

In this study, we used the eye tracker to record and analyze learners' visual attention distributions over a multimedia (PPT) presentation in a real classroom where an instructor gave a lecture on the topic of dinosaurs. Our study showed that during the presentation, the students allocated much of their attention to both the PPT presentation and to the instructor. When viewing the PPT slides, visual attention increased for those slides containing both text and picture components, compared to those with pictures alone. Among the text and picture zones (excluding the hypothesis slides), the text zones in general received higher percentages of viewing time. It was found that the average fixation duration, which relates strongly to the cognitive activities required to process information, was longer on pictures than on text. Among the text–picture slides, the slides on the teacher-initiated evidence obtained higher average fixation durations. When viewing the slides describing scientific hypotheses, the difference in visual attention between the text and picture zones reduced, except for the one slide illustrating the plate tectonics hypothesis. Further statistical analysis revealed that the effect of prior knowledge was evident mostly in the text zones. Although in most cases there was no difference in the viewing time of the pictures between the different background groups, further analyses of the densities of fixations revealed that the ES students knew better where to look. Finally, the analysis of the saccade paths showed that the inter-zone scanning, indicating integration of different modes of presentation, was evident during the PPT presentation. As expected, the ES students performed generally better than did the NES learners.

From these findings, several inferences are worthy of further discussion. First, the written texts seemed to be a major source of information for learning during the multimedia presentation. The crucial role of text in multimedia learning has been documented by some previous studies (e.g., Hegarty & Just, 1993; Liu & Chuang, 2011; Schmidt-Weigand et al., 2010). In these studies, the advantage of auditory text over written text is frequently reported. However in this study, although the instructor provided an auditory input of information that

Table 6
Group differences in numbers of saccade scanning between text and picture look-zones.

PPT page	Design	BK	SSP (TSP/TTT)	ITP (IBTP/TTT)
How can you know there were dinosaurs?	The first slide organizing the content (containing both text and picture)	ES	11.36** (0.89)	7.64** (0.63)
		NES	5.10 (1.00)	3.00 (0.65)
		Mean	8.38 (0.95)	5.43 (0.64)
Nest (S)	Text and picture	ES	11.64 (0.41)	8.36 (0.29)
		NES	8.40 (0.38)	5.60 (0.24)
		Mean	10.10 (0.40)	7.05 (0.27)
Fossil eggs (S)	Text and picture	ES	19.73* (0.49**)	14.55** (0.36**)
		NES	13.40 (0.34)	8.80 (0.22)
		Mean	16.71 (0.42)	11.81 (0.29)
Tooth marks	Text and picture	ES	13.82** (0.44)	8.91* (0.29)
		NES	10.10 (0.40)	6.20 (0.23)
		Mean	12.05 (0.41)	7.62 (0.26)
Teeth	Text and picture with the animated highlight	ES	9.00 (0.48*)	6.73* (0.34*)
		NES	6.30 (0.27)	3.30 (0.16)
		Mean	7.71 (0.38)	5.10 (0.29)
Skin marks	Text and picture with the static highlight	ES	15.18 (0.44)	12.82 (0.36**)
		NES	14.30 (0.37)	10.30 (0.25)
		Mean	14.76 (0.40)	11.62 (0.31)
Cause of extinction (Celestial cause)	Text and a conceptual graphic (planet model)	ES	7.55** (0.50)	6.18** (0.25)
		NES	3.30 (0.38)	2.60 (0.38)
		Mean	5.52 (0.44)	4.48 (0.31)
Cause of extinction (Plate tectonics)	Text and a conceptual graphic (Cartoon model)	ES	11.18 (0.64*)	9.18 (0.52*)
		NES	9.40 (0.49)	7.90 (0.41)
		Mean	10.33 (0.57)	8.57 (0.47)
Cause of extinction (Climate change)	Text and a numerical table	ES	7.82 (0.62)	5.45 (0.43)
		NES	6.00 (0.47)	4.50 (0.35)
		Mean	6.95 (0.55)	5.50 (0.39)

Note: (1) SSP: Sum of saccade paths; (2) ITP: Sum of the saccade paths between text and picture zones; (3) TSP/TTT = Total saccade paths/Total time tracked; (4) IBTP/TTT = Integration between text and picture/Total time tracked; (5) **$p < 0.05$; *$p < 0.1$.

was identical to the written text on the PPT slides, the students did not focus their attention more on the picture zone. Instead, their attention stayed largely on the text zones. This significant attention to the written texts was even more apparent among the ES students who had higher relevant knowledge and who also scored higher on the memory test.

In multimedia learning, many studies have shown the redundancy effect when auditory and written modes of presentation appear together (Kalyuga, 2009; Mayer, 2005). In this study, the instructor's narrations continued throughout the presentation. When the slides changed from photo-alone format to text–picture format, the students' overall visual attention increased. The analysis of viewing time indicated that much of the learners' attention was devoted to the written text zones. Meanwhile, as shown in Table 2, the percentages of viewing time suggested that the students still paid attention to the instructor's explanations. Accordingly, it could be concluded that their cognitive efforts did increase in the case of the text–picture slides in comparison with the picture-alone ones. However, whether the increased cognitive efforts for processing different modes in the verbal system hindered learning should be more carefully considered. In this study, it was found that the ES students attended more to the written text zones than did the NES students, and they recalled the concepts significantly better after the lecture. As a matter of fact, between written texts and pictures, the NES students also attended more to the text zones. In short, regardless of the students' background, the written text mode of information was preferred even though narration was provided.

The discussion above leads to the conclusion that both auditory and written modes of presentation in the real classroom helped the concept learning. It was likely that the narrations by the instructor helped the learners to activate their prior knowledge and locate the critical written information on the slides. This result is not exactly consistent with the modality effect in which auditory information is preferred to written information in multimedia presentations (Mayer, 2001), even though the individual differences have been widely discussed in recent studies (e.g., Mayer, 2005). It also cannot be explained satisfactorily by the redundancy effect that redundant information increases cognitive load and would hinder learning (e.g., Moreno & Mayer, 1999; Kalyuga, 2009). The findings in our study indicate that the two verbal modes of information could have been playing different roles during the multimedia learning in the real classroom. However, how exactly visual attention interacts with learning outcomes needs to be further examined.

Second, our study suggests that there is an interaction between types of graphic and information processing behaviors, and that this interaction is mediated by prior knowledge. As described above, when the students viewed slides describing the conceptual models, the difference in their visual attention distributions between the text and picture zones decreased. This effect was particularly evident for those slides regarding the asteroid impact and climate change. In other words, among the text–picture slides, attention to the conceptual graphics was higher than to the photos. The conceptual graphics are more complicated and abstract than actual photos; therefore, more attention to the conceptual graphics is conceivable. However, while the attention difference between the text and picture zones was reduced for the asteroid impact and the climate change slides, it was still significantly large for the "plate tectonics" slide. That is, the students' attention in average was focused more on the texts when reading the plate tectonics slide. Noticeably, the average fixation duration on the "plate tectonics" graphic was the highest, implying higher cognitive effort needed for processing the information presented in this particular graphic.

It is likely that since the theory of plate tectonics is a formal explanation for dinosaur extinction in the discipline of earth science, the NS students did not need to spend much time reading the graphic to understand the concept. Consequently, their attention remained largely on the text zone. As a result, the difference of visual attention between text and picture on the plate tectonics slide increased. On the other hand, since the "plate tectonics" hypothesis was less frequently heard than the other two hypotheses by the NES students, they had to attend more to the graphic in order to understand the "new" concept. The analysis indicated in Table 5 confirms that the NES students distributed a higher percentage of viewing time than did the ES individuals to inspect the graphic of plate tectonics. Accordingly, the high average fixation duration on the graphic, along with the higher need for graphic reading, could be an indicator of learning difficulty for that particular slide. In short, the above discussion points out that the interaction between types of graphics and information processing behaviors is also mediated by prior knowledge.

Third, our study supports that prior knowledge affects concept learning in the processes of both information decoding and integration. The analysis of the densities of fixations on the three hypothesis slides suggests that the ES students who had better concept gains knew better where to look. In addition, according to the analysis of the saccade paths, the ES students performed the integrative process more frequently. These findings are parallel to those of an eye study conducted by Humphrey and Underwood (2009) who employed a sophisticated scanpath comparison showing that domain knowledge mediates the visual saliency in scene recognition. The analyses of times of saccade paths and fixation density in the study imply that the influences of prior knowledge appeared in both the processes of information decoding and integration.

Finally, whether the auditory mode of information is preferred during multimedia learning may depend on its delivery source. Many previous studies of multimedia learning have shown the advantage of auditory over written presentation. However, in this study, with the instructor delivering the course content identical to the text descriptions on the PPT slides, the written texts, compared to the pictures, still attracted significant visual attention from the learners. We have argued previously that the narrations given by the instructor could have played a role in guiding visual attention to the written text. It is believed that narrations delivered by an instructor in a real classroom are by nature different from recorded narrations, as many experiments have found. Narration by the instructor, who is often regarded as an authority and a source of information, may give rise to different information processing behaviors. Such a conjecture is worthy of further examination.

6. Educational implications

Based on the results of this study, some educational implications can be drawn. First, when presenting multimedia material in the real classroom, a written text or note about course content is necessary for enhancing learners' visual attention. Second, although longer average fixation durations were expected to be found for pictures (Rayner, Smith, Malcolm, & Henderson, 2009), our data shows that the percentage of text reading time was high on slides with the conceptual graphics. This finding suggests that the conceptual graphics, due to their complexity, could have exerted extraneous cognitive load on the learners. As a result, the learners sought other information sources, i.e., the text, to help them comprehend the pictures. To reduce extraneous cognitive load, the use of the cuing strategy for pictures could be critical (e.g., Lin & Atkinson, 2011). Our study shows that the animated highlight seemed to be able to direct students' attention more to the picture, and reduced the percentage of viewing time on the text.

Third, given that the students with prior knowledge performed better in the processes of information decoding and integration, effective multimedia material should incorporate instructional strategies that can direct learners' attention to keywords, and allow learners to think and link information across different modes of presentation. These strategies could be especially important for novice learners. The use of different cuing strategies, such as organization signals, annotation and so forth, can help learners to locate important information in the text (e.g., Lai, Tsai, & Yu, 2011; Lorch & Lorch, 1996). To encourage information integration, the multimedia material may incorporate questions designed to stimulate learners' reflective thinking focusing on the linkage of different modes of presentation. The instructor may also post questions during the presentation allowing learners to link course information rather than merely narrating the course content. In short, the incorporation of instructional strategies in the multimedia materials will determine the effectiveness of multimedia presentations.

7. Research limitations and suggestions for future studies

It is necessary to note that since the study was conducted in an authentic classroom environment, and the eye-movement data were collected from one student at a time, there might be some dynamic factors that could reduce the power of the interpretations and inferences we made from the data, even though we made efforts to ensure that the students received the same quality of instruction. These factors may include the varying pace of narration delivered by the instructor, the variety of simultaneous interactions between the instructor and the participants, different questions raised by students during the lecture, student experiences and interests, and so forth. In short, naturalistic studies could be more uncertain than experimental studies, which already have a well-defined design tradition (Holmqvist et al., 2011). Despite these shortcomings, this study demonstrates an attempt to present students' online information processing behavior during a multimedia presentation. Naturalistic studies, we believe, though difficult to control well, can provide different perspectives for understanding authentic learning behaviors. We sincerely hope that more similar studies and analyses can be conducted in a variety of classroom settings so that gradually students' actual cognitive processes during learning can be disclosed.

In the study, we used the fixation density and the number of saccade paths as indices for the effect of prior knowledge. Though the two measures could show a general picture of how students of different backgrounds differ in the reading of the PPT information, they are limited in showing students' processing strategies. Many researchers who study visual or scene perception advocate the analysis of "scanpaths," i.e., the sequential patterns of eye fixations, to unveil people's visual search processes (Josephson & Holmes, 2002; Underwood, Humphrey, & Foulsham, 2008; Zangemeister, Sherman, & Stark, 1995). According to the scanpath theory, eye movements during imagery are stable, reversible and controlled by an internalized, cognitive perceptual model (Brandt & Stark, 1997). Accordingly, the analysis of scanpaths is expected to reveal how people memorize or comprehend visual information. In recent years, the analysis of scanpaths has also been applied to eye movement studies on text and symbolic reading to disclose how readers analyze and integrate syntactic or symbolic information (Jansen, Marriott, & Yelland, 2007; von der Malsburg & Vasishth, 2011). Thus, for future studies, scanpath analysis that reflects sequential patterns of cognitive control over attention is recommended to help depict how students of different academic backgrounds acquire subject knowledge.

In the study, we assessed students' concept gains by a free-recall memory test. The memory test indicated that the ES students seemed to be more capable of reproducing conceptual ideas introduced in the presentation. However, the result of the memory test could not be further inferred to questions of whether students of related background are better at constructing sophisticated mental models of the issue in discussion, or applying learned knowledge in problem solving. Different test formats, such as yes–no questions, essays, open-ended questions or conflicting tests/pictures, should be utilized in future studies to reveal different aspects of learners' conceptual achievements. A further comparison between eye movement data and concept achievements will reveal how different eye movement patterns affect information processing and retention.

Finally, the issue of the precision of the eye movement data is worth further discussion. The application of eye movement technology in the area of education has grown in recent years. Although the technology is able to provide a process view regarding how learners attend to and encode the learning materials, a fundamental question is whether the eye movement data do indeed reflect how the learner perceives as he/she is reading through the learning material. The reporting of data quality is important in that it indicates whether the research data can be trusted. However, in the literature, most of the eye movement studies, in particular in the areas of reading and education, report only the accuracy data. It seems that the precision issue is left to be determined by the developer. In fact, there is far from any consensus regarding what basic eye-movement information should be presented in the publications (Holmqvist, Nyström, & Mulvey, 2012).

In recent years, some scholars have started to advocate the importance of data quality. Detailed measurements of the precision errors and other data quality requirements can now be found in the literature (Holmqvist et al., 2011). However, different eye trackers with different fixation algorithms seem to have different considerations regarding the data quality issue. It could be more complicated when different research needs are taken into account. As educators whose primary aim is to study students' learning processes by using eye trackers, we expect that the measurements of data quality should be much easier to assess. Therefore, it is strongly recommended that the reporting of precision and other necessary quality information should be standardized and can be acquired automatically during data collection. To carry out such an implementation, more communication is needed among educators, eye movement theorists, researchers, and technology developers.

References

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*, 27–38.
Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169–200). NY: Cambridge University Press.

Change, C. C., Lei, H., & Tseng, J. S. (2011). Media presentation mode, English listening comprehension and cognitive load in ubiquitous learning environments: modality effect or redundancy effect? *Australasian Journal of Educational Technology, 27*, 633–654.

Dillon, A., & Jobus, J. (2005). Multimedia learning with hypermedia. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 569–588). Cambridge MA: Cambridge University Press.

Fletcher, L., Loy, G., Barnes, N., & Zelinsky, A. (2005). Correlating driver gaze with the road scene for driver assistance systems. *Robotics and Autonomous Systems, 52*, 71–84.

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: patterns in saccade direction vary with picture orientation. *Vision Research, 48*, 1777–1790.

Greene, J. A., Costa, L. J., Robertson, J., Pan, Y., & Deekens, V. M. (2010). Exploring relations among college students' prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment. *Computers & Education, 55*, 1027–1043.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717–742.

Hoffler, T. N., & Schwartz, R. N. (2011). Effects of pacing and cognitive style across dynamic and non-dynamic representations. *Computers & Education, 57*, 1716–1726.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking – A comprehensive guide to methods and measures.* NY: Oxford University Press.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Data quality: what it is and how to measure it. In *Proceedings of the 2012 symposium on eye-tracking research and applications* (pp. 45–52). ACM.

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: the role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology, 23*, 1215–1226.

Humphrey, K., & Underwood, G. (2009). *Domain knowledge moderates the influence of visual saliency in scene recognition.*

Humphrey, K., & Underwood, G. (2010). The potency of people in pictures: evidence from sequences of eye fixations. *Journal of Vision, 10*(10), 1–10, 19.

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*(2), 172–176.

Jansen, A. R., Marriott, K., & Yelland, G. W. (2007). Parsing of algebraic expressions by experienced users of mathematics. *European Journal of Cognitive Psychology, 19*, 286–320.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Got, T. (2010). In the eyes of the beholder: how experts and novice interpret dynamic stimuli. *Learning and Instruction, 20*, 146–154.

Josephson, S., & Holmes, M. E. (2002). Attention to repeated images on the World-Wide Web: another look at scanpath theory. *Behavior Research Methods, Instruments & Computers, 34*, 539–548.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review, 87*, 329–355.

Kalyuga, S. (2009). *Managing cognitive load in adaptive multimedia learning.* PA: IGI Global.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Lai, Y. S., Tsai, H. H., & Yu, P. T. (2011). Integrating annotations into a dual-slide PowerPoint Presentation for classroom learning. *Educational Technology & Society, 14*, 43–57.

Larsson, P. (2002). *Automatic visual behavior analysis.* Unpublished master dissertation. Sweden: Linköping University.

Leslie, K. C., Low, R., Jin, P., & Sweller, J. (2012). Redundancy and expertise reversal effects when using educational technology to learn primary school science. *Educational Technology Research and Development, 60*, 1–13.

Lin, L. J., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concept and processes. *Computers & Education, 56*, 650–658.

Liu, H. C., Andre, T., & Greenbowe, T. (2008). The impact of learner's prior knowledge on their use of chemistry computer simulations: a case study. *Journal of Science Education and Technology, 17*, 466–482.

Liu, H. C., & Chuang, H. H. (2011). An examination of cognitive processing of multimedia information based on reviewers' eye movements. *Interactive Learning Environments, 19*, 503–517.

Liu, H.-C., Lai, M.-L., & Chuang, H.-H. (2011). Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers' cognitive processes. *Computers in Human Behavior, 27*(6), 2410–2417.

Lorch, R. F., & Lorch, E. P. (1996). Effects of organizational signals on free recall of expository text. *Journal of Educational Psychology, 88*, 38–48.

von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language, 65*, 109–127.

Mayer, R. E. (1997). Multimedia learning: are we asking the right question? *Educational Psychologist, 32*(1), 1–19.

Mayer, R. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. (2005). *The Cambridge handbook of multimedia learning.* NY: Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*(8), 760–769.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401.

Mintzes, J., Wandersee, J. H., & Novak, J. D. (Eds.), (2000). *Assessing science understanding* (pp. 304–341). San Diego, CA: Academic Press.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movement during search. *Vision Research, 46*, 614–621.

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

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford. England: Oxford University Press.

Paivio, A. (2006). *Mind and its evolution: A dual coding theoretical interpretation.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of Verbal Learning & Verbal Behavior, 20*, 532–539.

Pastore, R. (2012). The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education, 58*, 641–651.

Plass, J. L., Moreno, R., & Brunken, R. (Eds.), (2010). *Cognitive load theory.* NY: Cambridge University Press.

Purnell, K. N., & Solman, R. T. (1991). The influence of technical illustrations on students' comprehension in geography. *Reading Research Quarterly, 26*(3), 277–299.

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

Rayner, K. (2009). The thirty fifth Sir Frederick Bartlett lecture: eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science, 20*, 6–10.

Reisslein, J., Seeling, P., & Reissein, M. (2005). Computer-based instruction on multimedia networking fundamentals: equational versus graphic representation. *IEEE Transactions on Education, 48*, 438–447.

Rieber, L. P. (2005). Multimedia leanring in games, simulations and microworlds. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 554–567). Cambridge MA: Cambridge University Press.

Rummer, R., Schweppe, J., Furstenberg, A., Seufert, T., & Brunken, R. (2010). Working memory interference during processing texts and pictures: implications for the explanation of the modality effect. *Applied Cognitive Psychology, 24*, 164–176.

Sadoski, M., & Willson, V. L. (2006). Effects of a theoretically-based large scale reading intervention in a multicultural urban school district. *American Educational Research Journal, 43*, 137–154.

Salthouse, T. A., & Ellis, C. I. (1980). Determinants of eye-fixation duration. *American Journal of Psychology, 93*, 207–234.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the eye tracking research and applications symposium* (pp. 71–78). New York: ACM Press.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). Explaining the modality and contiguity effects: new insights from investigating students' viewing behavior. *Applied Cognitive Psychology, 24*, 226–237.

Schneider, W., & Korkel, J. (1989). The knowledge base and text recall: evidence from a short term longitudinal study. *Contemporary Educational Psychology, 14*, 382–393.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *NeuroReport, 9*, 2195–2200.

She, H.-C., & Chen, Y.-Z. (2009). The impact of multimedia effect on science learning: evidence from eye movements. *Computers & Education, 53*(4), 1297–1307.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12*, 257–285.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*, 643–659.

Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2011). Visual attention for solving multiple-choice science problem: an eye-tracking analysis. *Computers & Education, 58*, 375–385.

Tsai, J. L., Yen, M. H., & Wang, C. A. (2005). Eye movement recording and the application in research of reading Chinese. *Research in Applied Psychology, 28*, 91–104.

Underwood, G., Foulsham, T., van Loon, E., & Underwood, J. (2005). Visual attention, visual saliency, and eye movements during the inspection of natural scenes. In J. Mira, & J. R. Alvarez (Eds.), *Artificial intelligence and knowledge engineering applications: A bioinspired research* (pp. 459–486). Berlin: Springer.

Underwood, G., Humphrey, K., & Foulsham, T. (2008). Knowledge-based patterns of remembering: eye movement scanpaths reflect domain experience. In *HCI and usability for education and work. Lecture notes in computer science, Vol. 5298/2008* (pp. 125–144). Berlin/Heidelberg: Springer.

Vaughan, J., & Graefe, T. M. (1977). Delay of stimulus presentation after the saccade in visual search. *Perception & Psychophysics, 22*, 201–205.

Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction, 4*, 25–42.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*, 345–376.

Yang, F. Y., & Chang, C. C. (2009). Examining high-school students' preferences toward learning environments, personal beliefs and concept learning in web-based contexts. *Computers & Education, 52*, 848–857.

Young, K., Mitsopoulos-Rubens, E., Rudin-Brown, C. M., & Lenne, M. G. (2012). The effects of using a portable music player on simulated driving performance and task-sharing strategies. *Applied Ergonomics, 43*, 738–746.

Zangemeister, W. H., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia, 33*, 1009–1025.