Interactive Learning Environments
Vol. 19, No. 5, December 2011, 503–517

Routledge
Taylor & Francis Group

An examination of cognitive processing of multimedia information based on viewers' eye movements

Han-Chin Liu[a] and Hsueh-Hua Chuang[b]*

[a]*Department of E-Learning Design and Management, National Chiayi University, Chiayi, Taiwan;* [b]*Center for Teacher Education and Institute of Education, National Sun Yat-sen University, Kaohsiung, Taiwan*

(*Received 28 July 2009; final version received 23 November 2009*)

This study utilized qualitative and quantitative designs and eye-tracking technology to understand how viewers process multimedia information. Eye movement data were collected from eight college students (non-science majors) while they were viewing web pages containing different types of text and illustrations depicting the mechanism of atmospheric pressure and the formation of sea and land winds. The results showed that participants' eyes were fixated more often on the text than on the illustrations. Breaking the instructional multimedia into small successive segments did not seem to increase the number of eye fixations on the illustrations. Participants alternated their eye fixations on related components of the illustrations while focusing on verbs or sentences representing the kinematics of the weather systems. Text seemed to provide more detailed explanations than pictures and therefore served as the main resource for understanding the content. Scan paths revealed that participants were likely to be distracted by decorative icons in the illustrations. The decorative icons also created a split-attention effect on students' cognitive processing. Eye-tracking technology was found to be a valuable tool for achieving the objectives of this study, and it provided insights into students' cognitive processes during multimedia learning. Discussion and suggestions for instructional designers and future studies are provided.

Keywords: eye movements; eye-tracking; information processing; cognitive load; multimedia learning

Introduction

Multimedia that use a variety of resources to present information through multiple sensory modalities is commonly believed to serve the purposes of education and entertainment (Schnotz & Lowe, 2003). However, research has also shown that simply presenting information in different representational formats does not necessarily promote significant understanding of the information in certain content domains (Mayer, Hegarty, Mayer, & Campbell, 2005; Moreno, 2004). Ploetzner and Lowe (2004) argued that researchers should take into account interactions between learners and multiple forms of representations to better understand how mental models are constructed for multimedia learning environments. Accordingly, Schnotz

*Corresponding author. Email: hsuehhua@gmail.com or hhchuang@mail.nsysu.edu.tw

ISSN 1049-4820 print/ISSN 1744-5191 online

http://dx.doi.org/10.1080/10494820903520123
http://www.tandfonline.com

and Lowe (2003) suggested that educators and researchers focused more on effects of multimedia on learning at the semiotic and sensory levels than at the technical level in multimedia learning environments.

If we are to fully understand how viewers process the information presented in learning materials, alternate methods for collecting objective, real-time data must be applied. One possibility is eye-tracking technology, through which we can analyze how learners interact with different forms of multimedia representations. Eye movement patterns provide clues about how information encoded in different visual media formats are retrieved and processed (Underwood & Radach, 1998). Although many studies on reading and image perception have shown effective utilization of eye-tracking technology in identifying cognitive processes, few have implemented such technology to uncover exactly how viewers process information that is encoded in a combination of text and illustrations (Jacob & Karn, 2003). In this study, we utilized eye-tracking technology to explore learners' information processing patterns when the information is encoded in different formats. By analyzing eye-movement data, we also sought to uncover how different forms of presentations affect learners' processing of information in multimedia learning environments.

Review of the relevant literature

Multimedia learning

New technologies, especially those involving computer-generated illustrations encoded in either a static or dynamic format, have been widely adopted by instructors and educators as supplements to traditional text-based learning materials. In 21st century classrooms, educational content is commonly encoded in multiple modalities. However, it is argued that it is the design of instruction, not simply the delivery of rich forms of media, that actually improves student learning (Clark, 1994). Ainsworth (2006) also argued that (a) the design of tasks and messages for multimedia learning, (b) support from multimedia in learning, and (c) the cognitive tasks required for interacting with multimedia, all must be analyzed if instructional effectiveness is to be determined. The so-called "media effect" found in early research was argued to be injudicious and conclusions drawn from the studies were somewhat doubtful because of overly simplistic comparisons of the effects produced by different media (Clark, 1994; Schnotz & Lowe, 2003).

Mayer (2001) defined multimedia as "the presentation of material using both pictures and words" (p. 2). Rather than focusing on how rich media are presented, Mayer (2001) proposed a "multimedia effect" that focuses on how the instructional multimedia effectively affects learners' cognitive processing of the content. Based on empirical results from several studies, Mayer also proposed a multimedia principle arguing that learning can be improved by using words and pictures rather than words alone. A vast number of studies have revealed the effectiveness of multimedia. These studies have also identified a variety of variables that might influence student learning, including individual's cognitive ability and style, the design of the learning task, and the integration of representations (Dutke & Rinck, 2006; Mayer & Massa, 2003; Moreno, 2004). Recent studies have also shifted attention from a delivery to a cognitive view of multimedia learning. How learning tasks are designed and the form of the multimedia message presented have both been found to play an important role in improving the cognitive processing of learning content (Hulshof, Eysink, & de Jong, 2006; Tabbers, Martens, & van Merriënboer, 2004). The cognitive tools provided by multimedia can support learning tasks for a wide range of individuals (Bera & Liu, 2006). Bera and Liu (2006) also investigated the users' patterns of use of cognitive tools in the problem space of multimedia learning. Their findings suggested that a more contextual approach to the cognitive processes benefits students' learning.

When employing different representational formats, learners must deal with diverse sources of information. Therefore, some of their mental effort may be invested in processing information encoded in different formats or representations that are unrelated to the content, rather than in deeper thinking tasks such as analyzing and organizing the content knowledge. Accordingly, those who seek to achieve the full potential of multimedia in helping students understand a particular content domain must take into account the students' cognitive tasks for information processing.

Cognitive processes in multimedia learning

The processing of messages encoded in multiple representations can be cognitively demanding. In a multimedia environment, the learner must identify information encoded in different representations and make sense of a plethora of incoming messages. Mayer (2005) argued that not all multimedia presentations are equally effective in promoting learning and proposed a cognitive theory of multimedia learning. Three assumptions serve as the background of Mayer's theory: dual-channel, limited capacity, and active processing. According to these assumptions, words are received through the verbal channel while pictures are received through the visual channel into the learner's working memory. Yet, words can be encoded in either the text or the narration format and can therefore be processed using either of the channels depending on the format. In addition, information is being actively selected, analyzed, and organized in the learner's working memory in preparation for storage into long-term memory. However, each individual has limited working memory capacity so an excessive amount of information may possibly increase the learner's cognitive workload and thus reduce information-processing efficiency. Poorly designed multimedia representations can require significant cognitive effort to process information unrelated to the intended content. Mayer's principles for multimedia learning, based on the framework of cognitive information processing, provide guidelines for designing multimedia messages for more effective learning.

Although on-screen text and illustrations are still the preferred way to present information in most multimedia learning environments, the relationship between text and graphics is still the focus of many multimedia studies (Paas, Touvinen, Tabbers, & Van Gerven, 2003; Sweller, Van Merriënboer, & Paas, 1998). Sweller (1994) argued that separating illustrations and the related explanatory text can reduce search efficiency and cause the so-called ''split-attention effect''. Therefore, Sweller suggested placing the illustrations close to related text in order to ameliorate this problem and conserve the cognitive resources for meaning making. Likewise, Mayer (2001) proposed the contiguity principle, suggesting that placing illustrations and related text instructions together increases not only the spatial contiguity but also the temporal contiguity, of the representations. Therefore, related information can be closely processed in one's working memory for better schema building. Additionally, Mayer (2001) suggested that splitting a large amount of information into smaller successive segments helps reduce the amount of incoming information to be processed, effectively conserving the limited working-memory capacity.

Most multimedia researchers have tested their hypotheses by measuring student understanding of content with comprehension-related test items. Little has been done to directly identify the learners' cognitive activities to investigate the impact of information modality on student learning, and students' real-time cognitive processes have rarely been tracked in a non-intrusive way. Pass and van Merriënboer (1993) used paper-and-pencil surveys to measuring learners' cognitive load when learners were engaged in learning activities. Another approach, which also could have the undesirable effect of increasing mental workload, is the use of think-aloud methods that require students to describe their mental activities during learning activities. Such subjective and retrospective procedures for tracking cognitive processes can damage the validity of the findings, because they rely heavily on respondents' memory and awareness. This therefore calls for alternative means, other than subjective and retrospective measures, to provide evidence in support of various principles of multimedia design intended to promote learning. Mayer (2001, 2005) recommended that more research on the direct tracking of cognitive processes during learning tasks is needed if we are to identify the concepts necessary to develop valid theories of multimedia learning. Such learning requires the cognitive processing of different representations of incoming information. The design of effective multimedia instruction would be facilitated if educators and researchers could understand how different individuals cognitively utilize multimedia information. A research design that employs alternative methodology should be beneficial in discovering insights into cognitive activities and strategies with which learners explore information encoded in multiple forms in multimedia learning.

Eye-tracking cognitive processes

Addressing information on a computer screen is essential for an individual using a computer to learn. The trace of an individual's eye movements can work as a blueprint that presents details as to how information encoded in different visual media formats is retrieved and processed (Underwood & Radach, 1998). Eye-tracking technology has been used by psychologists since the early 20^{th} century (Judd, McAllister, & Steel, 1905). With the technological advances in the late 1990s, eye-tracking has become easier to implement and more eye movement studies have emerged since then. This has allowed researchers to switch their attention from improving the technology to exploring human–computer interactions and related cognitive processes (Jacob & Karn, 2003).

An individual's eye movements reveal a continual stream of information that can be used to indicate her or his mental state. The positions and the number of fixations, the fixation duration, and the saccade length are the most common variables in eye-movement measures. Human eyes are believed to be stable only for short periods of roughly 200–300 ms and covering about 2° of visual angle in which the information is focused and is clear. The periods of stability of one's eyes are called fixations. Saccades refer to the fast and scattered movements of one's eyes from one fixation point to another. It is believed that no information is processed during such movements. The distance between two successive fixation points is defined as the saccade length. Eye movement variables and combinations of these variables can also be used as indicators of an individual's mental effort. Eye fixations are found to correspond to the information to be encoded and processed by the cognitive system. The correlation between fixation and mental processing of information is referred as the "eye–mind" assumption (Just & Carpenter, 1976; Rozenblit, Spivey, & Wojslawowicz, 1998). The duration of eye fixations, the number of fixations, and the amount of refixations reveal patterns describing how a user's attention is directed to a given region or visual area of the computer screen.

Jacob and Karn (2003) reviewed early studies and listed the metrics of the eye movement variables that are most commonly used to measure and identify the associated cognitive activity. They claimed that the number of fixations is correlated with the efficiency of searching for information on a computer screen; the more fixations, the less efficient the search. The duration that a viewer gazes at a particular component of the visual can help in identifying the viewer's area of interest. The duration of fixations can also mirror the viewer's difficulty level encountered during information processing. The viewer's area of interest can also be identified using the frequency of fixations on a specific element or area of a visual display. Furthermore, the sequence of fixations or the scan paths illustrating the changes of the areas of interest over time can be used to map the viewer's mental imagery constructed during viewing (Huber & Krist, 2004).

Collecting eye movement data today is less intrusive compared to measuring other physiological quantities such as heart rate or electroencephalographs. Eye movement measures therefore appear to be a promising and minimally intrusive alternative for gathering real-time information about a learner's cognitive activities. Although the literature reveals that eye movement data have a strong potential for determining a viewer's cognitive activity, most studies have collected such data solely for reading or picture perception (Jacob & Karn, 2003). Most of today's printed and multimedia instructional materials offer a mix of text, pictorial, and auditory information to help students acquire understanding. Some studies have employed eye-tracking technology to investigate the perception of information encoded in combinations of visual and verbal formats (Hegarty, 1992; Hegarty & Just, 1993; Stolk, Boon, & Smulders, 1993). For example, some eye-tracking studies have found that viewers usually inspect pictorial information after they finish reading the text (Carroll, Young, & Guertin, 1992; Hegarty & Just, 1993; Rayner, Rotello, Stewart, Keir, & Duffy, 2001). Hegarty and Just (1993) utilized eye-tracking technology to explore readers' eye movement patterns while they were reading text accompanied by illustrations. In their study, the eye fixation patterns revealed that the readers relied mainly on the text as the information source, although the illustrations did help them build mental models of particular pulley systems.

Only a few multimedia learning studies have focused on tracking viewers' eye-movement patterns. Thus, the improvement of eye-tracking technology has increased the opportunities for in-depth and less-intrusive investigations of information processing in multimedia learning environments. In addition to providing comprehensive test and self-reporting survey data, eye-tracking technology appears to be a valuable tool for the objective assessment of learners' mental effort and cognitive processes, thereby providing a common basis for educators, researchers, and multimedia designers to improve multimedia learning.

The present study

In the present study, we applied eye-movement measures to determine how learners process information in multimedia learning environments. By collecting these eye-movement data, we intended to explore how learners process the information encoded in different formats, such as text and illustrations. By collecting information on learners' cognitive processes, we also sought to understand in depth how the participants interacted with the multimedia learning materials and to provide suggestions for effective designs of instructional multimedia.

Methodology

Participants and design

The design of the study has both quantitative and qualitative elements, with participants' eye movement patterns and numbers of eye fixations as the main data sources. Eye movements were traced by eye-tracking equipment that provided qualitative patterns and quantitative data for an in-depth analysis of how people process information on a computer screen. Eight non-science majors in their freshman or sophomore years of college voluntarily participated in the study.

Materials

Three sets of web-based multimedia presentations were developed for this study, covering topics of formation of land winds, sea winds, and high and low atmospheric pressures. The land wind presentation (LWP) consisted of five successive pages illustrating the step-by-step formation of land winds (see Figure 1). The text explanation and the illustrations for the LWP was broken into five successive portions presented on the five web pages to depict step-by-step the formation of land wind phenomenon. The dimension of the illustration on each of the land wind page is 6.61×3.40 in., taking up an area of 22.47 in.^{2} of the computer screen with an 800×600 pixel resolution. The dimension of the text area on each land wind page is different because of the different amount of words. As a result, the total text area of the land wind pages is 26.52 in.^{2} while the total illustration area is 112.37 in.^{2} (see Figure 1).

The sea wind presentation (SWP) contained a single page of illustrations and text demonstrating the formation of sea winds (see Figure 2). The illustrations on the sea

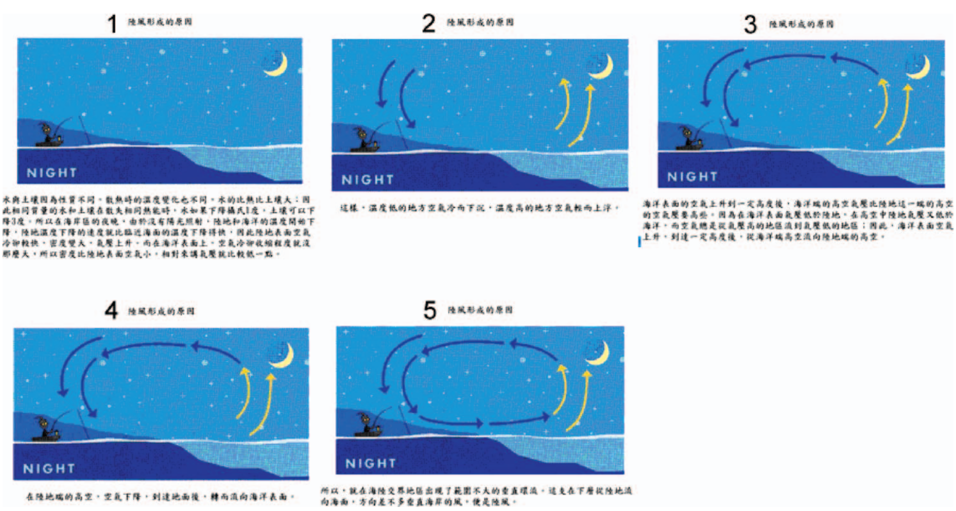

Figure 1. Land wind series pages.

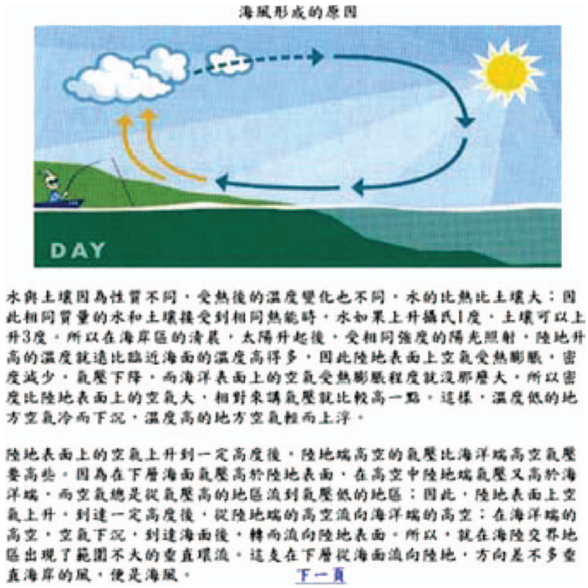

Figure 2. Sea wind page.

wind page were 6.34 in. × 3.08 in., taking up an area of 19.53 in.^{2} on the computer screen, whereas the text area was 7.39 in. × 3.44 in., taking up an area of 25.42 in.^{2}. The resolution was 800 × 600 pixels.

Except for reversal air movements, the formations of land and sea winds share a common mechanism; they both are caused by atmospheric pressure differences at the land and sea areas. Therefore, the numbers of words in the text explanations were modified to be the same in the LWP and the SWP. The same structure was used in the atmospheric pressure presentation (APP) as in the SWP, but the text explanation for the APP was shorter (see Figure 3). The three illustrations in the APP were three-dimensional, as we wanted to know if this stylistic change would have an impact on cognitive processing. The three APP illustrations were 4.82 × 2.28, 2.28 × 1.40, and 2.28 × 1.40, respectively, taking up a total area of 17.37 in.^{2}. The corresponding text area was 2.90 × 4.49 in., taking up 13.02 in^2. The resolution again was 800 × 600 pixels.

Equipment

A faceLAB™ 4 eye-tracking system was used to collect real-time eye movement data. This system includes a computer workstation, a set of eye-tracking cameras, an infrared tracker, and a personal computer that runs the instructional multimedia. While the information on the computer screen is being retrieved, the eye-tracking system uses desktop cameras to non-intrusively track the viewer's eye movements, including fixations, fixation durations, saccades, scan path, and pupil dilations.

Procedure

The order of presentation of the three page types (APP, LWP, and SWP) was randomized across participants. The eye-tracking system was calibrated before each

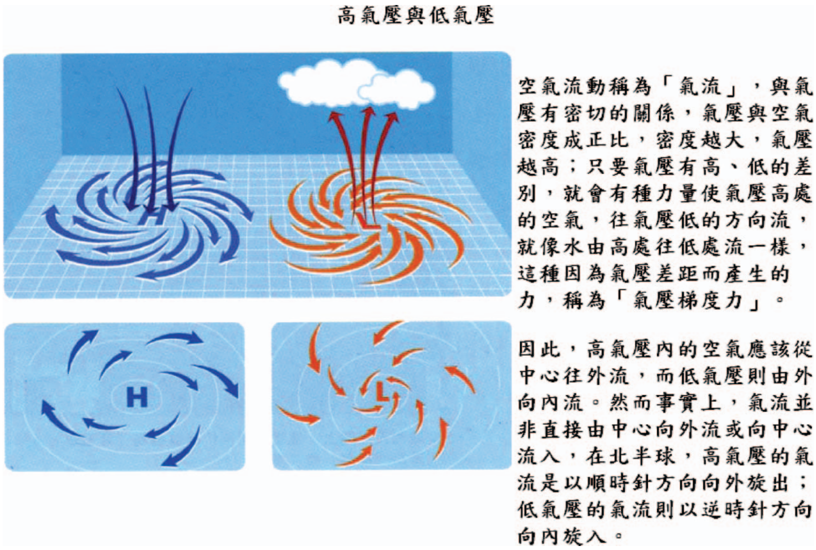

Figure 3. The "Formation of atmospheric pressure" page.

experimental session began. The participants spent roughly 5–10 min prior to each trial to allow the system to adjust to tracking their eye-pupil positions.

Data analysis

Data on the numbers of eye fixations on both the illustration and text areas on each page were collected for analysis. Because the total viewing areas differed for the three presentations, participants' total number and duration of eye fixations were divided by the sizes of the representations to derive measures respectively called fixation density and duration density. These measures allowed us to compare participants' attention to the different presentations in either medium (illustrations or texts). The scan paths and fixation positions were treated as a blueprint of the information processing. These information processing patterns and strategies were compared with one another to determine if differences in the forms of the media presentations had an impact on participants' cognitive processing.

Results and discussion

Eye-movement patterns for text and illustrations

The average number of fixations, fixation density, duration, and duration density for text and illustrations are shown in Tables 1 and 2 as a function of information type and instructional content. The numbers and durations of fixations were summed across the five land-wind pages to obtain overall values for the two areas of the LWP. Table 1 and 2 show that, except for the LWP pages, the participants fixated their eyes more frequently on text than on illustrations: for APP, $F(1, 7) = 9.84$, $p < 0.001$, $MSE = 4.42$; for LWP, $F(1, 7) = 1.66$, $p = 0.14$, $MSE = 26.44$; for SWP, $F(1, 7) = 4.17$, $p = 0.001$, $MSE = 26.87$. Also, the participants spent significantly more time viewing text than viewing illustrations: for APP, $F(1, 7) = 9.78$, $p < 0.001$, $MSE = 1.32$; for LWP, $F(1, 7) = 1.89$, $p = 0.10$, $MSE = 7.62$;

Table 1. Mean number of fixations and fixation density (fixations per square of inches) on different web page presentations.

	APP, mean (SD)	LWP, mean (SD)	SWP, mean (SD)
Number of fixations			
Text	56.88 (22.93)	100.63 (44.44)	120.38 (75.65)
Illustration	13.38 (15.26)	56.63 (43.67)	16.63 (9.71)
Fixation density			
Text	4.37 (1.76)	3.79 (1.68)	4.74 (2.98)
Illustration	0.77 (0.88)	0.50 (0.39)	0.85 (0.50)

Table 2. Mean durations (in seconds) and duration density (seconds per square of inches) on different web page presentations.

	APP, mean (SD)	LWP, mean (SD)	SWP, mean (SD)
Duration			
Text	17.26 (7.00)	30.90 (13.78)	36.83 (24.25)
Illustration	4.34 (4.72)	16.51 (11.62)	5.00 (2.92)
Duration density			
Text	1.33 (.54)	1.17 (.52)	1.45 (.95)
Illustration	0.25 (0.27)	0.15 (0.10)	0.26 (0.15)

for SWP, $F(1, 7) = 3.99$, $p = 0.005$, $MSE = 7.98$. Likewise, when the viewing areas of the individual presentations were taken into consideration, participants were found to fixate more often on text than on illustrations: for APP, $F(1, 7) = 9.25$, $p < 0.001$, $MSE = 0.39$; for SWP, $F(1, 7) = 4.95$, $p = 0.002$, $MSE = 0.66$; for LWP, $F(1, 7) = 4.05$, $p = 0.005$, $MSE = 0.96$. They also spent significantly more time (duration) viewing text than viewing illustrations: for APP, $F(1, 7) = 9.20$, $p < 0.001$, $MSE = 0.12$; for LWP, $F(1, 7) = 5.04$, $p = 0.002$, $MSE = 0.20$; for SWP, $F(1, 7) = 3.88$, $p = 0.006$, $MSE = 0.31$.

The division of the LWP material into small segments was intended to conserve working memory space during the processing of the illustrations. However, given the relatively large amount of pictorial information on the LWP pages, reducing the number of words in each LWP segment did not seem to direct participants' attention more frequently, or for a longer time, to the illustrations.

In short, our results confirm the argument that when text and illustrations are presented on the same screen, users spend more time on the text than on the illustrations (Hegarty, 1992; Rayner et al., 2001). In other words, we found that on-screen text messages seemed to dominate the model-building process. Our conclusion is also confirmed at the individual level. Figure 4 provides fixation distributions for three cases in which the SWP page was observed. Both the quantitative and qualitative findings show that the participants' attention was mainly text-directed.

Integration of text and pictorial information

Hegarty and Just (1993) found that their participants tended to integrate text and pictorial information either locally (by switching their eyes between text and illustration) or globally (by gazing at most if not all of the components of the illustration, after reading the text, to build mental models). In this study, when total

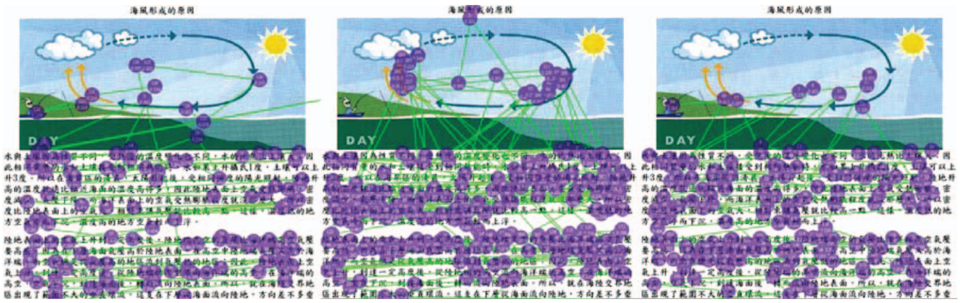

Figure 4. Sample fixation distributions on the sea wind page.

number of saccades (summed across the three presentations) between the illustration and the end of each sentence were compared to the number of saccades between the illustration and other parts of each sentence, no significant differences were found, $t = 1.31$, $p = 0.23$. However, when the numbers of local and global inspections of each of the three presentations were further analyzed, we found that participants made significantly more global than local inspections only on the APP pages: for APP, $F(1, 7) = 3.97$, $p = 0.005$, $MSE = 0.378$; for SWP, $F(1, 7) = 0.114$, $p = 0.91$, $MSE = 1.09$; for LWP, $F(1, 7) = 1.31$, $p = 0.23$, $MSE = 0.67$. Note that the APP contained fewer words than the other two presentations, and also that the illustrations on the APP pages were three-dimensional and more detailed than the two-dimensional illustrations on the LWP and SWP pages (see Figures 1–3). This greater detail in the APP illustrations made it more likely that participants would inspect the text and illustrations more globally to construct their mental models of the atmospheric pressure systems.

Participants were especially likely to refer to the visuals when there was text describing the kinematics of the system (see Figure 5). Specifically, in such circumstances, the fixations on the illustrations were short and focused on what had just been read. Carroll et al. (1992) argued that viewers' attention to the pictorial and text information in particular media is isolated and that they usually look at a picture after they finish reading the text. Rayner et al. (2001) found that viewers usually read through text messages before looking at the illustrations, rather than shift back and forth between text and illustrations. Echoing the above-mentioned findings, the scan paths in our data reveal more horizontal regressions of the fixations on the text, with few vertical fixation shifts between text and illustrations (see Figures 6 and 7). Arrows were used to represent air movement. When eye fixations reached particular adjectives in the text, such as "clockwise" and "inward" (indicating the movement of the air), the next eye fixation was commonly found to switch to the arrows in the illustrations (see Figure 6). Figure 6 also shows a scan path in which the participant shifted her fixation from the word "clockwise" (circled and numbered "2" in the text) to an illustration of low atmospheric pressure after she read the word "inward" (circled and numbered "1") in the text. The participant then fixated her eyes on the illustration of high atmospheric pressure (circled and numbered "4") after having attended to the word "clockwise" (circled and numbered "3").

The majority of participants' saccades were in the text area, with a limited number of fixation switches between text and illustrations (see Figure 6). This result suggests that viewers utilized the text information to construct their mental models,

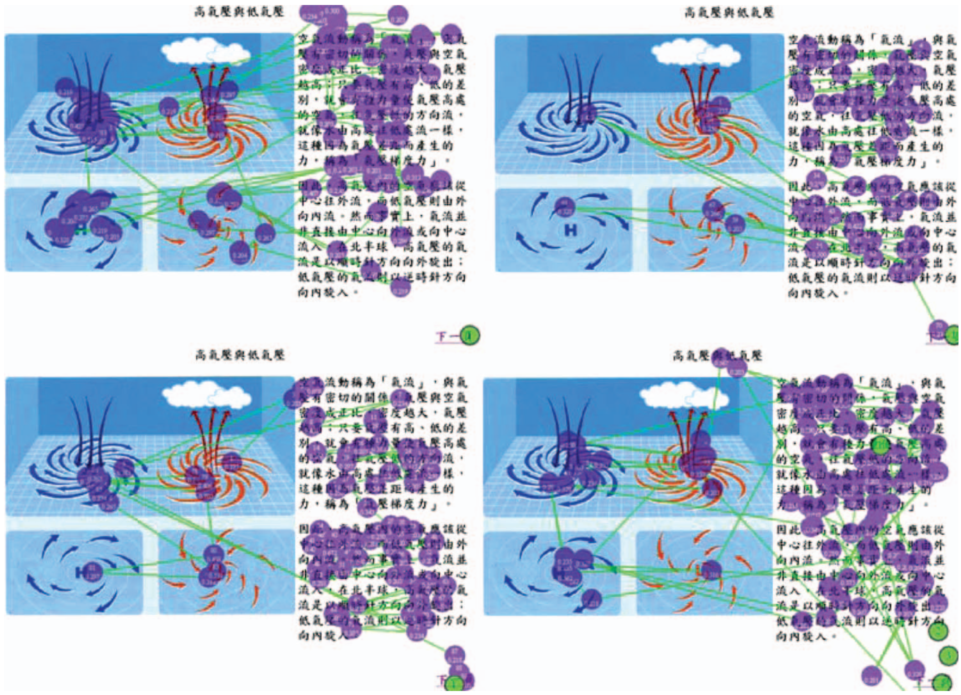

Figure 5. Sample fixation distributions on the atmospheric pressure page.

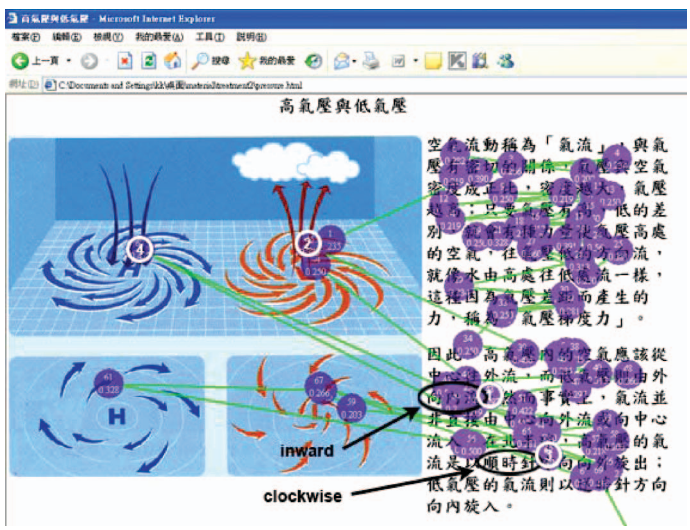

Figure 6. A representative scan path on the atmospheric pressure web page.

perhaps because the text information was more detailed. Mayer (2001) argued that viewers mentally convert selected on-screen text into a "sound image" to form a verbal model of the representation. To make sense of the multimedia information, the verbal model is then integrated with the pictorial model that had previously been constructed from perceived images and prior knowledge. However, others have argued that illustrations commonly served as external supplements to the memory of a text description to construct a mental model of the system (Hegarty, 1992; Hegarty & Just, 1993). In the present study, the eye fixations clustered on particular areas of the text and illustrations. Mayer (2001) argued that for multimedia learning, people must select specific words and images in verbal and visual channels, respectively, because of the limited capacity of verbal and visual working memory. The findings in our study suggest that viewers pay attention only to particular locations in the text and illustrations because of limited working-memory capacity.

The effect of redundant images

Another interesting finding concerns the attractiveness of decorative images. An icon of a fisherman fishing was located in the lower left corners of illustrations on both the sea wind and the land wind pages (see Figure 7). The fisherman icon served as a decoration; it was not related to the information content.

Some of the scan paths indicate that several fixations were directed at the fisherman icon when the participant was viewing the text (see Figure 8). Early studies showed that viewers' processing of pictorial information seemed to be directed by the

Figure 7. Sample illustrations of the redundant icons.

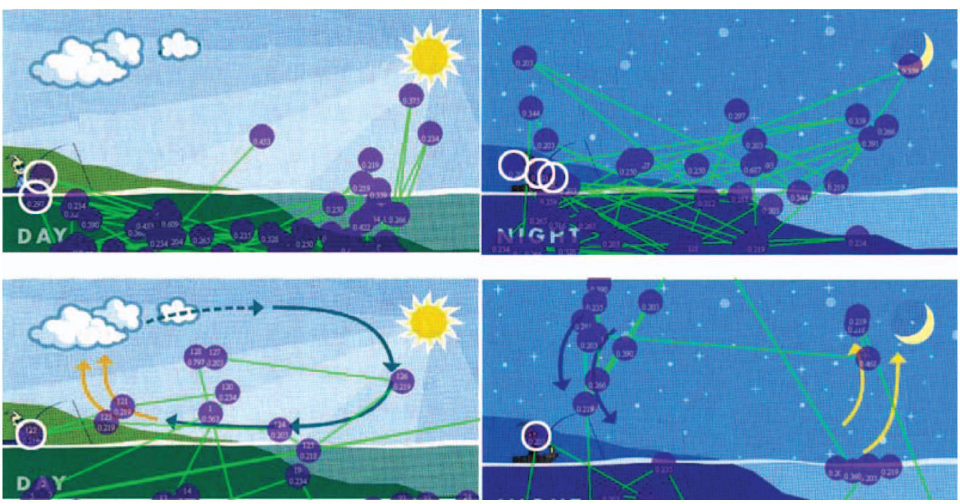

Figure 8. Representative fixations on redundant images.

text information (Caroll et al., 1992; Hegarty & Just, 1993; Rayner et al., 2001). As the viewers' saccades between text and illustrations were argued to demonstrate the viewers' strategies in processing multimedia information for later construction of mental models (Hegarty & Just, 1993), the fixations on the decorative icons in saccades are likely to interrupt the processing of multimedia information.

According to Mayer (2001), interesting but topic-irrelevant images could work as distractions that draw students' attention away from the target representations of the learning content. In other words, adding interesting but irrelevant icons and images can increase cognitive load and thus inhibit learning. The fixations on the fisherman icon in our study suggest that the decorative images indeed attracted participants' attention and therefore could have made it more difficult for them to process the target information.

Conclusion

Text and illustrations continue to be the major media in multimedia learning; therefore, they must be carefully crafted to effectively enable understanding. In this study, the scan paths and distributions of the participants' eye fixations showed that when illustrations and related text were grouped together, participants tended to utilize the text as the main vehicle for understanding the content to be learned, because the text generally provided more detail. Breaking the text information into small successive segments did not seem to promote participants' awareness of the pictorial information. The nature of the illustrations also could have impacted how successfully the participants processed the multimedia information. Participants referred to the more detailed illustrations most often after they had read particular verbs and sentences that were related to the components of the pictorial information. Echoing previous studies, we also found that the main approach of our participants to building a mental model of a particular weather system was first to read and reread the text pertaining to the kinematics, for configuration purposes; afterwards, they often redirected their attention to the pictorial information. These findings suggest that the pictorial information that designers of instructional multimedia provide should be detailed, especially when static illustrations are used to supplement text explanations of the functions of particular systems. Also, our participants sometimes appeared to be attracted to specific decorative images in the illustrations. Instructional multimedia designers should thus minimize decorative images to avoid drawing viewers' attention away from processing the primary information to be learned.

Although the impact of multimedia on learners' understanding of various topics has been found to be significant, only a handful of practitioners have utilized alternatives other than achievement tests to achieve an in-depth understanding of how learners process information. Although previous researchers have employed subjective and retrospective methods to investigate the effectiveness of multimedia learning, future researchers could take advantage of eye-tracking technology to create a more comprehensive understanding of how learners make use of the information presented by a rich combination of media.

Acknowledgment

This research project was supported in part by the National Science Council, Taiwan; Grant no: NSC-97-2511-S-415-005-MY2.

Notes on contributors

Dr. Han-Chin Liu is an assistant professor in the Department of E-Learning Design and Management at National Chiayi University in Taiwan. His research interests include technology integrated learning and multimedia learning.

Dr. Hsueh-Hua Chuang is an associate professor in the Center for Teacher Education and Institute of Education at National Sun Yat-Sen University in Taiwan. Her research interests include online learning, teacher adoption of technology, and multimedia learning.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*, 183–198.

Bera, S., & Liu, M. (2006). Cognitive tools, individual differences, and group processing as mediating factors in a hypermedia environment. *Computers in Human Behavior*, *22*, 295–319.

Carroll, P.J., Young, J.R., & Guertin, M.S. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 444–461). New York: Springer-Verlag.

Clark, R.E. (1994). Media will never influence learning. *Educational Technology Research and Development*, *42*, 21–29.

Dutke, S., & Rinck, M. (2006). Multimedia learning: Working memory and the learning of word and picture diagrams. *Learning and Instruction*, *16*, 526–537.

Hegarty, M. (1992). The mechanics of comprehension and comprehension of mechanics. In K. Rayner (Ed.), *Eye movement and visual cognition: Scene perception and reading* (pp. 428–443). New York: Springer-Verlag.

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

Huber, S., & Krist, H. (2004). When is the ball going to hit the ground? Duration estimates, eye movements, and mental imagery of object motion. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 431.

Hulshof, C., Eysink, T., & de Jong, T. (2006). The ZAP project: Designing interactive computer tools for learning psychology. *Innovations in Education and Teaching International*, *43*, 337–351.

Jacob, R.J.K., & Karn, K.S. (2003). Eye-tracking in human-computer interaction and usability research: Ready to deliver the promises [Section commentary]. In J.R. Hyona & H.D. Radach (Eds.), *The mind's eyes: Cognitive and applied aspects of eye movements*. Oxford, UK: Elsevier Science.

Judd, C.H., McAllister, C.N., & Steel, W.M. (1905). General introduction to a series of studies of eye movements by means of kinetoscopic photographs. In J.M. Baldwin, H.C. Warren, & C.H. Judd (Eds.), *Psychological review, monograph supplements* (Vol. 7, pp. 1–16). Baltimore: Review Publishing Company.

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

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

Mayer, R.E. (2005). Cognitive theory of multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R.E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology-Applied*, *11*, 256–265.

Mayer, R.E., & Massa, L.J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology*, *95*, 833–846.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*, 99–113.

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P.W.M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71.

Pass, F., & van Merriënboer, J.J.G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors*, *35*, 737–743.

Ploetzner, R., & Lowe, R.K. (2004). Dynamic visualizations and learning. *Learning and Instruction*, *14*, 235–240.

Rayner, K., Rotello, C.M., Stewart, A.J., Keir, J., & Duffy, S.A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, *7*, 219–226.

Rozenblit, L., Spivey, M., & Wojslawowicz, J. (1998, August). *Mechanical reasoning about gear-and-beltdiagrams: Do eye-movements predict performance?* Paper presented at the meeting of the Mind III: The Annual Conference of the Cognitive Science Society of Ireland, Dublin City University, Ireland.

Schnotz, W., & Lowe, R. (2003). External and internal representations in multimedia learning: Introduction. *Learning and Instruction*, *13*, 117–123.

Stolk, H., Boon, K., & Smulders, M. (1993). Visual information processing in a study task using text and pictures. In G. d'Ydewalle & J. van Rensbergen (Eds.), *Perception and cognition* (pp. 285–296). Amsterdam: Elsevier Science.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312.

Sweller, J., van Merriënboer, J.J.G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.

Tabbers, H.K., Martens, R.L., & van Merrienboer, J.J.G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, *74*, 71–81.

Underwood, G., & Radach, R. (1998). Eye guidance and visual information processing: Reading, visual search, picture perception and driving. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 1–28). Oxford, UK: Elsevier Science.