Computers & Education 58 (2012) 375–385

Contents lists available at ScienceDirect

Computers & Education

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

Visual attention for solving multiple-choice science problem: An eye-tracking analysis

Meng-Jung Tsai [a,*], Huei-Tse Hou [b], Meng-Lung Lai [c], Wan-Yi Liu [a], Fang-Ying Yang [d]

[a] Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, #43, Sec. 4, Keelung Rd., Taipei 106, Taiwan, ROC
[b] Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taiwan, ROC
[c] The Department of Early Childhood Education, National Chiayi University, Taiwan, ROC
[d] Graduate Institute of Science Education, National Taiwan Normal University, Taiwan, ROC

A R T I C L E I N F O

Article history:
Received 30 October 2010
Received in revised form
20 July 2011
Accepted 23 July 2011

Keywords:
Teaching/learning strategies
Applications in subject areas
Media in education
Interdisciplinary projects
Evaluation methodology

A B S T R A C T

This study employed an eye-tracking technique to examine students' visual attention when solving a multiple-choice science problem. Six university students participated in a problem-solving task to predict occurrences of landslide hazards from four images representing four combinations of four factors. Participants' responses and visual attention were recorded by an eye tracker. Participants were asked to think aloud during the entire task. A 4 (options) $\times$ 4 (factors) repeated measures design, two paired *t*-tests and effect sizes analyses were conducted to compare the fixation duration between chosen and rejected options and between relevant and irrelevant factors. Content analyses were performed to analyze participants' responses and think aloud protocols and to examine individual's Hot Zone image. Finally, sequential analysis on fixated LookZones was further utilized to compare the scan patterns between successful and unsuccessful problem solvers. The results showed that, while solving an image-based multiple-choice science problem, students, in general, paid more attention to chosen options than rejected alternatives, and spent more time inspecting relevant factors than irrelevant ones. Additionally, successful problem solvers focused more on relevant factors, while unsuccessful problem solvers experienced difficulties in decoding the problem, in recognizing the relevant factors, and in self-regulating of concentration. Future study can be done to examine the reliability and the usability of providing adaptive instructional scaffoldings for problem solving according to students' visual attention allocations and transformations in a larger scale. Eye-tracking techniques are suggested to be used to deeply explore the cognitive process during e-learning and be applied to future online assessment systems.

1. Introduction

The eye tracking technique has been typically adopted to examine human visual attention based on the eye-mind assumption (Just & Carpenter, 1980). In general, eye fixation location reflects attention and eye fixation duration reflects processing difficulty and amount of attention (the longer the information is fixated, the more complex it is or the deeper it is processed). Specifically, fixation duration varies on types of information (e.g. texts or graphics) and types of tasks (e.g. reading or problem solving). Furthermore, fixation locations and duration reflect the individuals' reading strategies and prior knowledge or experience (Hyönä, Lorch, & Kaakinen, 2002). Besides, scan path patterns exhibit individuals' cognitive strategies utilized in goal-oriented tasks (Gandini, Lemaire, & Dufau, 2008).

The eye-tracking method has been successfully applied in research fields including reading (Paulson & Jenry, 2002; Rayner, Chace, Slattery, & Ashby, 2006) and information processing (for a detailed review, see Rayner, 1998; Radach & Kennedy, 2004), arithmetic problem solving (Hegarty, Mayer, & Green, 1992; Verschaffel, De Corte, & Pauwels, 1992), human–computer interactions (Jacob & Karn, 2003) and emergent literacy (Evans & Saint-Aubin, 2005; Justice, Skibbe, Canning, & Lankford, 2005). Within these studies, we are particularly interested in how students solve problems in science.

* Corresponding author. Tel./fax: +886 2 27376270.
E-mail addresses: mjtsai99@mail.ntust.edu.tw, mjtsai99@gmail.com (M.-J. Tsai).

0360-1315/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2011.07.012

Previous eye-tracking studies focused on mathematics problem solving. Hegarty et al. have conducted several eye-tracking studies to examine the comprehension process and the strategies for solving mathematics word problems (Hegarty et al., 1992; Hegarty, Mayer, & Monk, 1995). They found that key information such as numbers and variable names to solving problems were fixated longer and were critical to the construction of the solution. In addition, compared to low accuracy students, high-accuracy students spent more time on more difficult problems and this extra amount of time was located in the integration and planning phases of problem solving. They also suggested that high-accuracy students made more use of the problem-model strategy (e.g., focusing more on variable names), and that low accuracy students made more use of the direct-translation strategy (e.g., focusing more on the numbers and relational words such as more or less). Verschaffel et al. (1992) also used eye-tracking technique to examine students' word problem solving and indicated that students made more comprehension reversal errors (e.g., addition used while subtraction was the correct strategy) when the order of the terms in the relational statement is not consistent with the preferred order. The above studies suggested that recognizing, selecting, and processing the relevant information is essential for successful mathematics word problem solving.

However, the paucity of studies on problem solving in science education is very surprising given its practical and theoretical importance. Recently, Tai, Loehr, and Brigham (2006) utilized the eye-tracking technique to explore what students with different subject backgrounds (chemistry, biology and physics) looked at while solving the standardized science examination problems. They found that the more expertise that the novices had in the specific subject, the fewer the eye fixations needed to process information in specific zones (e.g., problem statement zone, graph or image zone, multiple choices zone) and the fewer the saccades (abrupt, rapid eye-gaze movement between different positions) between zones. Although the above study employed images in problem solving tasks, it emphasized on the impact of prior knowledge on visual attention. Limited research has been reported on how students solve image- or graphic- based problems with multiple choices, which is frequently seen in science assessments. Our study extend prior research on science problem solving in two directions. First, we test how students solve problems in earth science. Second, we examine how students inspect complex graphics used in problem solving context.

Indeed, recently, the image-based eye-tracking studies have drawn interests from educational researchers due to the eye-tracking technology is increasingly user-friendly (e.g., more application packages are available for data analyses). Pilot studies have attempted to use eye-tracking technology to explore students' computer game learning experience (Alkan & Cagiltay, 2007), examine the multimedia effect on science learning (She & Chen, 2009), re-examine multimedia learning theories or enhance multimedia learning (for a special issue, see van Gog & Scheiter, 2010), and discuss its unique contributions in the study of multimedia learning (Hyönä, 2010; Mayer, 2010). However, little studies of multimedia learning which used eye-tracking technology was situated in the problem solving context. For example, Cook, Wiebe, and Carter (2008) observed students' visual attention on conceptually relevant features when interpreting multiple graphic representations. Jarodzka, Scheiter, Gerjets and van Gog (2010) examined students' visual attention while perceiving and interpreting complex and dynamic visual stimuli (such as videos). Finally, Canham and Hegarty (2010) explored students' attention allocations during the process of comprehension of complex graphics (such as weather maps). All of the above three studies focused on the prior knowledge variable and reported that students' prior knowledge had an impact on their visual attention allocations within a multimedia instructional context.

In sum, eye-tracking studies thus far have provided insights on how students pay attention to image-based learning materials in some instructional contexts. This study further explores how students inspect the images in a problem solving context.

2. Purpose

This study aimed at examining students' visual attention while solving a science problem with multiple choices using the eye-tracking technique. Specifically, this study explored how students spend time inspecting the *options* (i.e., items for answer selections, for example, A, B, C and D) and *factors* (i.e., icons for conceptual representations, for example, Temperature, Rainfall, Slope and Debris) embedded in an image-based problem and whether successful and unsuccessful problem solvers inspected the images in a different manner. Research questions were twofold:

1. How do students spend time inspecting *options* and *factors* in a science problem with multiple choices? Do they spend more time inspecting the *chosen* options compared to the *rejected* alternatives? Do they spend more time inspecting the *relevant* factors compared to the *irrelevant* factors?
2. How do *successful* and *unsuccessful* problem solvers inspect the *options* and *factors* while solving this science problem with multiple choices? Do they inspect the options and factors with different amount of attention? Do they inspect the options and factors in different sequences?

To answer these two research questions, an exploratory study with eye-tracking technique was adopted to observe students' eye-movements, especially focusing on their attention (i.e. fixation durations) on specific areas (i.e., LookZones) and the sequence patterns of their fixated LookZones. Prior eye-tracking studies have provided an adequate amount of evidences demonstrating that people tend to have longer fixations on more important or complex information (Hyönä, 2010; Rayner, 1998) and pay more attention to relevant information than irrelevnt information (Kaakinan, Hyönä and Keenan, 2002). A recent study (Jarodzka et al., 2010) further showed that the experts attended more to relevant aspects of dynamic multimedia stimuli compared to the novices. Based on the aforementioned literature, four hypotheses were proposed:

H1. Students, in general, spent more time inspecting their *chosen* options than *rejected* ones;

H2. Students, in general, spent more time inspecting *relevant* factors than *irrelevant* ones;

H3. Successful problem solvers inspected the *options* in a different pattern from unsuccessful problem solvers;

H4. Successful problem solvers inspected the *factors* in a different pattern from unsuccessful problem solvers.

3. Method

3.1. Participants and design

Six male university students, aged from 19 to 21, participated in this study. All of them majored in computer engineering and took fundamental earth science courses in high schools. Therefore, they already possessed some prior knowledge for solving the science problem. All participants had good visions and passed the eye-tracking calibrations. Four images in which four factors (Temperature, Rainfall, Slope and Debris) were included were designed to be inspected by each participant during the problem solving task.

3.2. Materials

A science problem regarding predictions of debris slide hazards was used for experiment in this study (see Fig. 1), in which four images representing four combinations of four possible factors. The materials were developed by a research group including one earth science expert and checked by another earth science expert outside the group before testing. The images were shown on a computer screen. On the top of the screen, the problem was initiated by a statement describing, "*Please select the image(s) inferring a landslide would occur (single/multiple selection) and justify your selection(s)*." The four possible factors included *three relevant factors* (rainfall density [Rainfall], slide slope [Slope], and debris on a slide [Debris]) and *one irrelevant factor* (temperature [Temperature]), with the latter especially designed for detecting possible individual differences in distinguishing or selecting relevant information for problem solving. Four options were embedded in each image (labeled from A to D) representing four combinations of the four factors. To successfully solve this problem, participants needed to distinguish relevant factors and induce an occurrence of landslide based on adequate extent of each factor. The combinations of these factors were the major keys to solving this problem. The best option was B because option B contained all required conditions causing a landslide, which were a rainfall density greater than 35 mm/h, a land slope larger than 15° and less than 35°, and with mass debris on a slide. However, option A could also be a correct answer, if the trees were shallow-rooted and therefore the debris could flow (and/or be transported) along with the trees during heavy rains. This given situation frequently occurs in some serious debris-flow hazard areas. Therefore, option A along with option B, i.e. AB, was also regarded as a correct answer if participants orally justified so for selection. Next to each image was a checkbox for participants to select this option. Participants were allowed to check more than one checkbox.

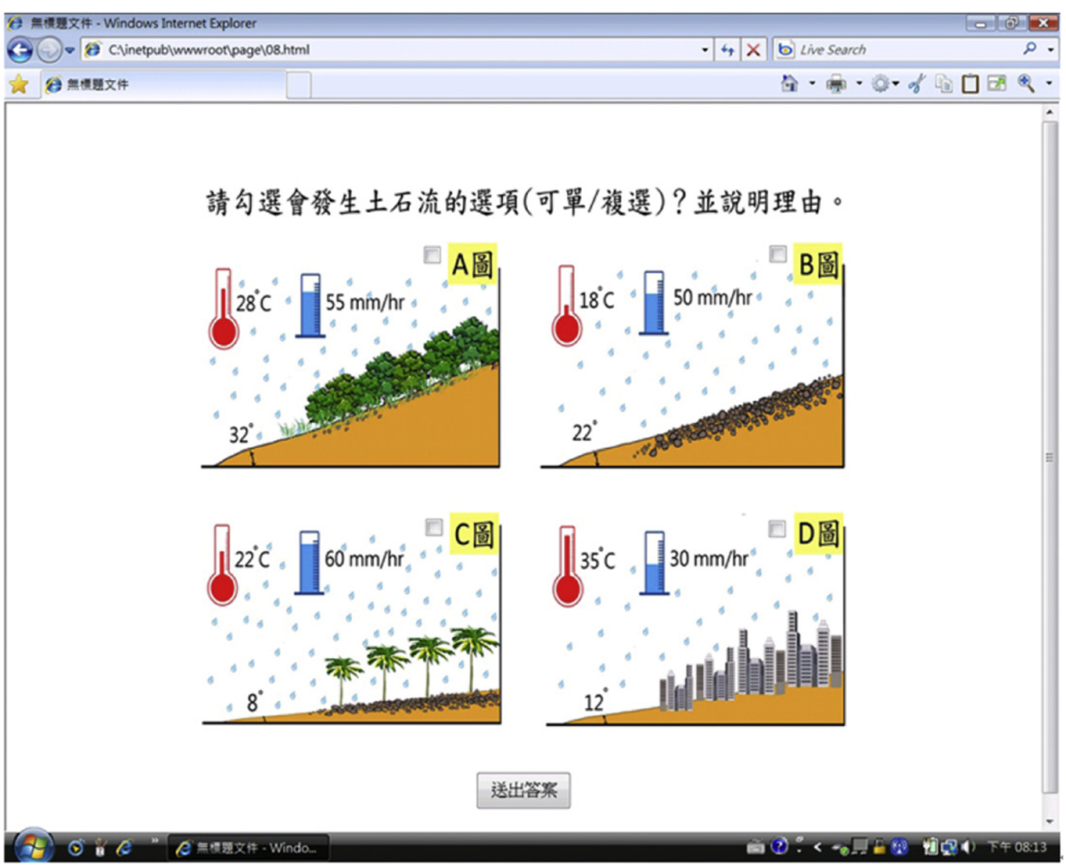

Fig. 1. The multiple-choice problem with four images embedding four factors used in the problem solving task. Participants were instructed by the title stating, "*Please select the image(s) inferring a landslide would occur (single/multiple selection) and justify your selection(s)*." A checkbox was provided aside each option and a submit button was provided in the bottom of the screen.

3.3. Apparatus

An eye tracker (faceLAB 4.5) with a sampling rate of 60 Hz was used to record participants' eye-movements. Fixations occurred when a subject's gaze stabilized over at least 200 ms (Duchowski, 2002). While collecting data, participants had free head and eye movement. The eye tracker was interfaced and controlled via a desktop, and another laptop was used to control the experiment. A 22-inch flat panel monitor with a resolution of 1280 × 1024 was used to display the experimental stimuli to the participants. The monitor was placed approximately 78 cm from the eyes of the participants. Participants responded to stimuli using a mouse and a keyboard. GazeTracker 8.0 software installed in the laptop was used to store and analyze the gaze data.

3.4. Procedure

All of the participants passed the calibrations with an accepted angular error of less than 1.0° measured by the eye tracker faceLAB 4.5. Therefore, all of them were included in the eye-tracking experiment to solve the science problem regarding predictions of landslide. In addition, participants were asked to think aloud while solving the problem. By doing so, we could accumulate participants' justifications which were used to check against their selection(s). A think aloud training was conducted before the experiment started. The entire experiment lasted approximately 30 min. Participants' eye movement data, question responses, and computer events such as typing and mouse clicking were all recorded by GazeTracker 8.0. Besides, participants' verbal reports of think aloud were audiotaped during the entire experiment.

3.5. Data analysis

3.5.1. Coding for LookZones

Before analyzing the data, several LookZones (see the rectangles in Fig. 2) were defined as follows. LookZone "Title" referred to the area of problem statement. For options, LookZone A, B, C and D indicated the entire area of image A, B, C and D, respectively. For factors, LookZone T1, T2, T3 and T4 indicated the area representing the Temperature factor in image A, B, C and D respectively. The same scheme was also applied to the LookZones regarding Rainfall, Slope and Debris factors in all images, e.g. R1, R2, R3 and R4 referred to the Rainfall factor in option A, B, C and D respectively. Therefore, there were total 21 LookZones defined for data analyses in this study.

In addition, two coding schemes were used to label the LookZones defined in this study. For the LookZones of options, each image A, B, C and D was labeled either "chosen" or "rejected" according to each participant's responses; for example, participant #4 selected B as an

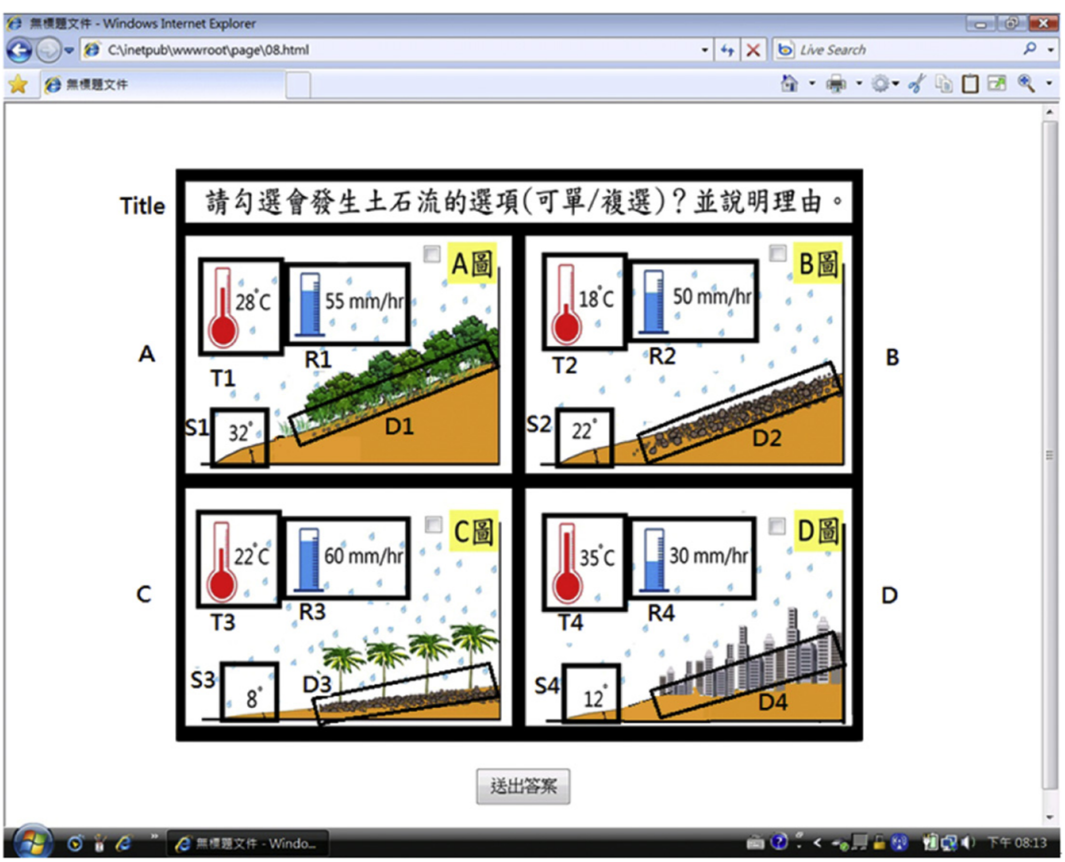

Fig. 2. The definition scheme for LookZones: LookZone *Title* was the problem statement, LookZone A, B, C and D indicated four options; and LookZone T, R, S and D represented Temperature, Rainfall, Slope and Debris factors respectively. The Chinese title states that "*Please select the image(s) inferring a landslide would occur (single/multiple selection) and justify your selection(s).*"

answer, so the LookZone B was labeled as "chosen" and LookZones A, C, and D were labeled as "rejected" for participant #4. As for the LookZones of factors, each T, R, S, and D was labeled as "relevant" or "irrelevant". Among the four factors, TRSD, T (temperature) is the only irrelevant one, as which does not cause landslide. Therefore, LookZones representing R (Rainfall), S (Slope), and D (Debris) were labeled as "relevant" while LookZones representing T were labeled as "irrelevant".

3.5.2. Approaches for data analyses

Both quantitative and qualitative approaches were used to answer the research questions. First, a 4 (options) $\times$ 4 (factors) repeated measures design was used to explore overall participants' fixation duration among all options and all factors. Two paired *t*-tests with effect size analyses were further used to compare the fixation durations between chosen and rejected options (for Hypothesis 1) and to compare the fixation durations between relevant and irrelevant factors (for Hypothesis 2).

Second, content analyses were conducted to examine participants' problem solving performances based on their responses to this problem and verbal reports (justifications). The responses and the transcripts of their verbal records were analyzed by two science educators with all agreements after discussions. Following the content analyses, an empirical observation on individual Hot Zone map (an image output of fixation allocations overlaid with a background image of the problem) was conducted for exploring possible group patterns. The Hot Zone images were created using MATLAB programming based on the characteristics of the human vision system (Wandell, 1995).

Moreover, sequential analysis method can explore the sequential relationships between learners' overall learning behaviors (Bakeman & Gottman, 1997; Jeong, 2003; Hou, 2010) (e.g., what kind of behavior follows after another certain type of behavior? Does this behavior sequence reach statistical significance?). Hence, a sequential analysis was further conducted to compare successful and unsuccessful problem solvers' inspecting sequences among fixated options (for Hypothesis 3) and among fixated factors (for Hypothesis 4).

4. Results

4.1. Fixation duration on options and factors

4.1.1. ANOVA with repeated measures design

To investigate whether participants preferred certain options and factors shown in the problem, we conducted a 4 (options) $\times$ 4 (factors) repeated measures ANOVA with participants' fixation duration on LookZones as the dependent measure. Because participants varied in the responding time (ranging from 64 to 132 s), in order to reduce the inter-subject variability (Cohen, Cohen, West, & Aiken, 2003), the fixation duration on different LookZones for each participant was represented as a proportion of his/her total fixation duration on the screen (i.e., the fixation duration on the LookZone/total fixation duration). Therefore, for each participant across all LookZones, *percentage of fixation duration* was calculated and analyzed.

The results of repeated measures analysis on percentage of fixation durations are shown in Table 1. A significant difference was found in fixation duration on *factors* but not *options*. No significant interaction was found. The post-hoc comparisons indicated that R received more attention than T, D, and S ($p = .015$, .023 and .001); also T and D received more attention than S ($ps < .05$). This demonstrated that, overall, the participants spent most of their time inspecting the Rainfall factor, then spent about equal time inspecting the Temperature and Debris factors, and spent least time inspecting the Slope factor. Fig. 3 further illustrated the percentage fixation duration distributed among all options and all factors.

4.1.2. Paired t-tests and effect sizes

To further examine hypotheses 1 and 2, two paired *t*-tests were conducted to compare the fixation duration between participants' chosen and rejected options, and also between relevant and irrelevant factors.

Regarding options, the paired *t*-test on *fixation duration* did not show any significant difference ($t = 1.995$, $df = 5$, $p = .103$), while the paired *t*-test on *percentage of fixation durations* between chosen ($M = 13.06$ s, $SD = 5.38$) and rejected ($M = 8.77$ s, $SD = 3.14$) showed a marginal significant difference ($t = 2.302$, $df = 5$, $p = .070$) between relevant ($M = .584$, $SD = .107$) and irrelevant ($M = .353$, $SD = .144$) factors. However, due to the small sample size, effect sizes were further calculated for investigation. The effect sizes of Cohen's *d* were .97 between chosen and rejected options and 1.83 between relevant and irrelevant factors, both reaching a high level of effect (see Fig. 4 for details). Based on the above, although paired *t*-tests just showed a marginal difference on fixation duration between relevant and irrelevant factors, effect sizes for both comparisons have reached practically significant levels for explanations. Therefore, the results seemed to support Hypothesis 1 and Hypothesis 2. That is, overall, participants tended to spend more time inspecting chosen options than rejected ones; and tended to spend more time inspecting relevant factors than irrelevant ones.

4.2. Problem solving performance

A content analysis of students' responses and their think aloud protocols was used to determine levels of problem solving performance. Table 2 summarized the responses of all six participants. Their reasons why selecting particular responses were also summarized from the

Table 1
ANOVA with 4 (option) $\times$ 4 (factor) repeated measures design plus post-hoc comparisons on percentage of fixation durations.

Source	SS	*df*	MS	*F*	*p*
Option (ABCD)	.001	3	.000	1.769	.196
Factor (RTSD)	.005	3	.002	14.679	.000***
Option $\times$ Factor	.001	9	.000	0.806	.613
Post Hoc (LSD)	R > T*; R > S**; R > D*; T > S *; D > S *; therefore R > T = D > S				

*$p < .05$; **$p < .01$; ***$p < .001$; T = Temperature; R = Rainfall; S=Slope; D = Debris.

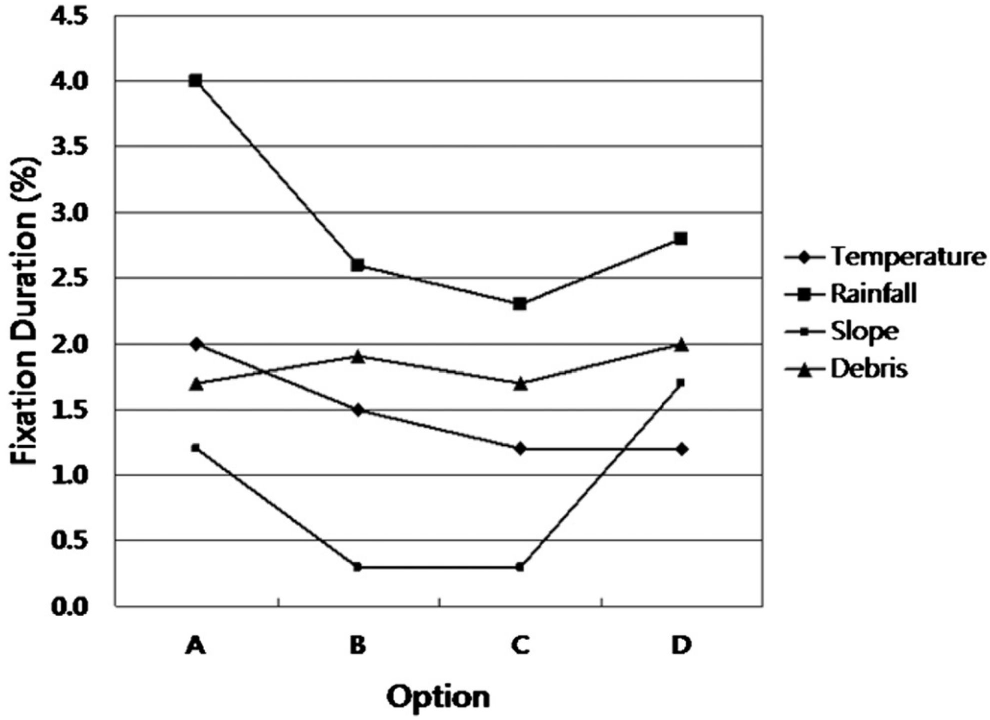

Fig. 3. Percentage of fixation duration allocated on the four factors (T, R, S and D) across the four options (A, B, C and D). Participants, in general, did not pay any different amount of attention among options but among factors (R > D = T > S, *ps*<.05).

transcripts of think aloud protocols. It was clear that participants #2, #4, and #5 selected B, while participants #1, #3, and #6 selected C, BC, and D, respectively. Because the correct answer was B or AB, the correct responses were coded as "Yes" for participants #2, #4, and #5 and as "No" for participants #1, #3, and #6. Regarding the justification for selections, numbers of relevant factors reported properly by each participant was also counted and recorded in Table 2. For example, participant #4 stated that, "*The reasons why I chose B was because the slide angle is large enough and there are no trees but rocks on the slide in image B*." This participant properly explained two relevant factors that cause a landslide (slope and debris); therefore, this response was coded as "2(S, D)" for participant #4. On the contrary, participant #6 responded, "*I think D should be the answer because it is in a city and the amount of the rainfall is normal*." "City" was not a relevant factor and "Rainfall" was not inferred properly for the problem as normal rainfall does not cause a landslide. Although Rainfall was mentioned but not explained properly, this response was coded as "0(R, not properly)" for participant #6. Based on the correctness of responses and numbers of relevant factors reported properly, the six participants were finally divided into two groups: high-score group (participants #2, #4, and #5) and low-score group (participants #1, #3, and #6). This grouping was used for further investigations in this study.

According to Table 2, it was clear that students performed differently in the multiple-choice problem solving and provided with different reasons for responses. An interesting observation was that, in the records of the relevant factors reported properly by each participant, the frequencies of the factors mentioned by participants showed an order of R (by 4 participants) > D (by 3 participants) > S (by 1 participant). Except for the Temperature factor, this result was quite consistent with the findings in the above post-hoc comparison of repeated measures regarding fixation duration allocations among factors (see Table 1). This indicated that students did not explicitly report their attention paid

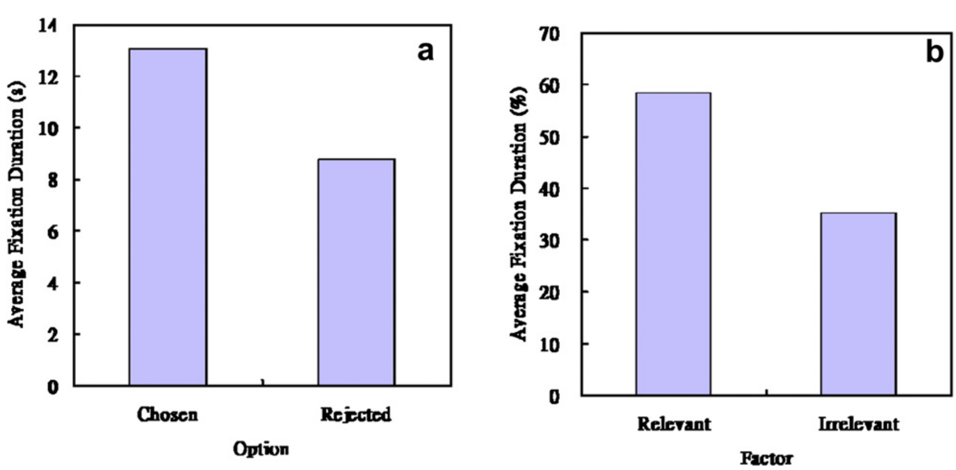

Fig. 4. Average fixation duration compared (a) between chosen and rejected options (Cohen's $d = 0.97$) and (b) between relevant and irrelevant factors (Cohen's $d = 1.83$).

Table 2
Performance evaluation of students' multiple-choice problem solving.

Subject ID	Answer	Reasons from verbal reporting	Correctness (B/AB: Yes; otherwise: No)	# of relevant factor reported properly (factors[a])	Level of performance
#1	C	"…because of the *heavy rainfall* and the arecas are not good for water and soil conservation."	No	1(R)	Low
#2	B	"…because of the *heavy rainfall* and *no trees but rocks anywhere.*"	Yes	2(R, D)	High
#3	BC	"…because no trees in B and arecas are not good for water and soil conservation in C…"	No	0(none)	Low
#4	B	"…because the *slide angle is large enough* and *no trees but rocks* on the slide."	Yes	2(S, D)	High
#5	B	"..because of the *higher rainfall* and *no trees but rocks* on the slide."	Yes	2(R, D)	High
#6	D	"…because it is in a city with a *normal rainfall*…"	No	0(R, not properly)	Low

[a] R = Rainfall, D = Debris, S=Slope.

on Temperature, the only irrelevant factor for this problem, which merits a further examination for explanations. Now that different problem solving performances were observed among the participants, we further explored the group differences by carefully looking into Hot Zone image for each participant.

4.3. Hot Zone images

In order to describe an entire picture of a participant's visual attention on the stimuli image, we analyzed the Hot Zone images (shown in Fig. 5 and Fig. 6) produced by MATLAB programming. In the image, the mapped color varied from individuals' total fixation duration at each pixel on the screen. The darker the red color was, the longer the total fixation duration accumulated on a pixel. Therefore, the red or orange spots (i.e., Hot Zones) represented locations where the information had been processed longer and more deeply by participants while blue colors represented locations where the information had been minimally processed.

By examining each participant's Hot Zone image, one interesting thing was found that individual differences existed in students' attention allocations among factors and options. For example, participant #4 (Fig. 5) and participant #6 (Fig. 6) paid most attention to options AB and AD, respectively, while participant #4 and participant #6 selected option B and D respectively. This was, to some degree, consistent with the results of the post-hoc comparison between chosen and rejected options, suggesting that while solving a multiple-

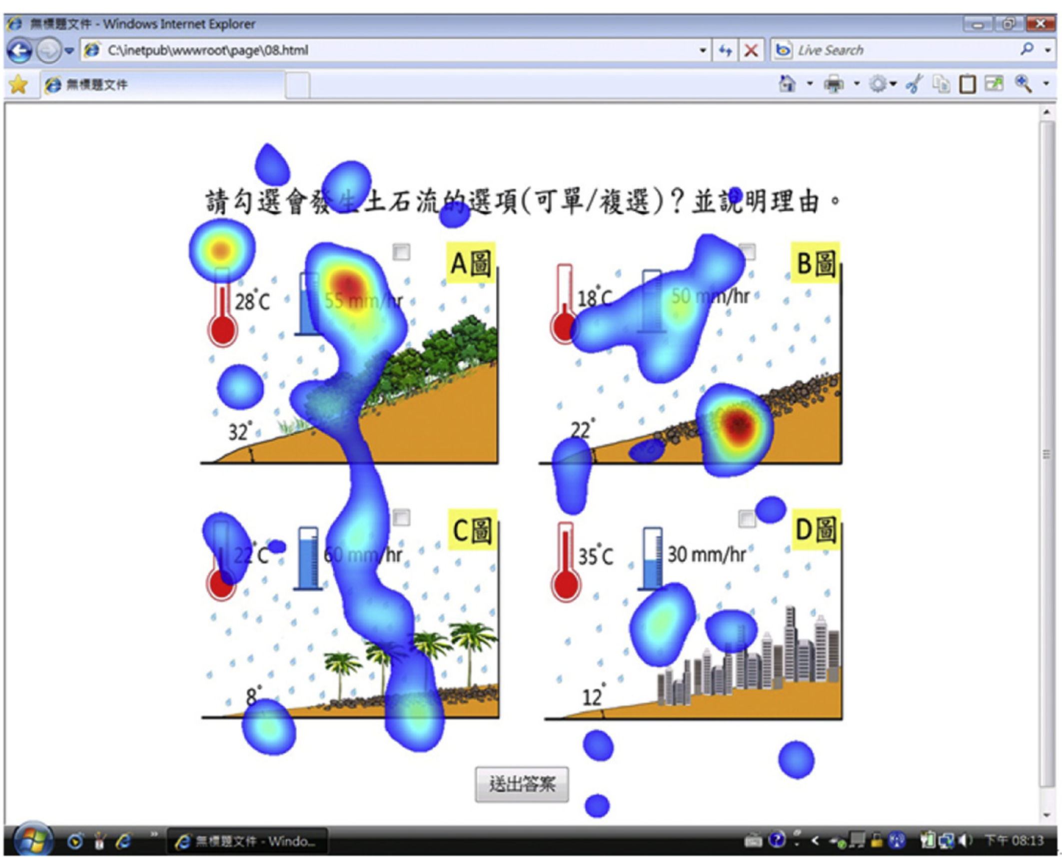

Fig. 5. The Hot Zone image for participant #4 (categorized in high-score performance group, with a response of B) showed that the participant paid most attention on option A and B, especially on factor R, D and T. The Chinese title states that "*Please select the image(s) inferring a landslide would occur (single/multiple selection) and justify your selection(s).*"

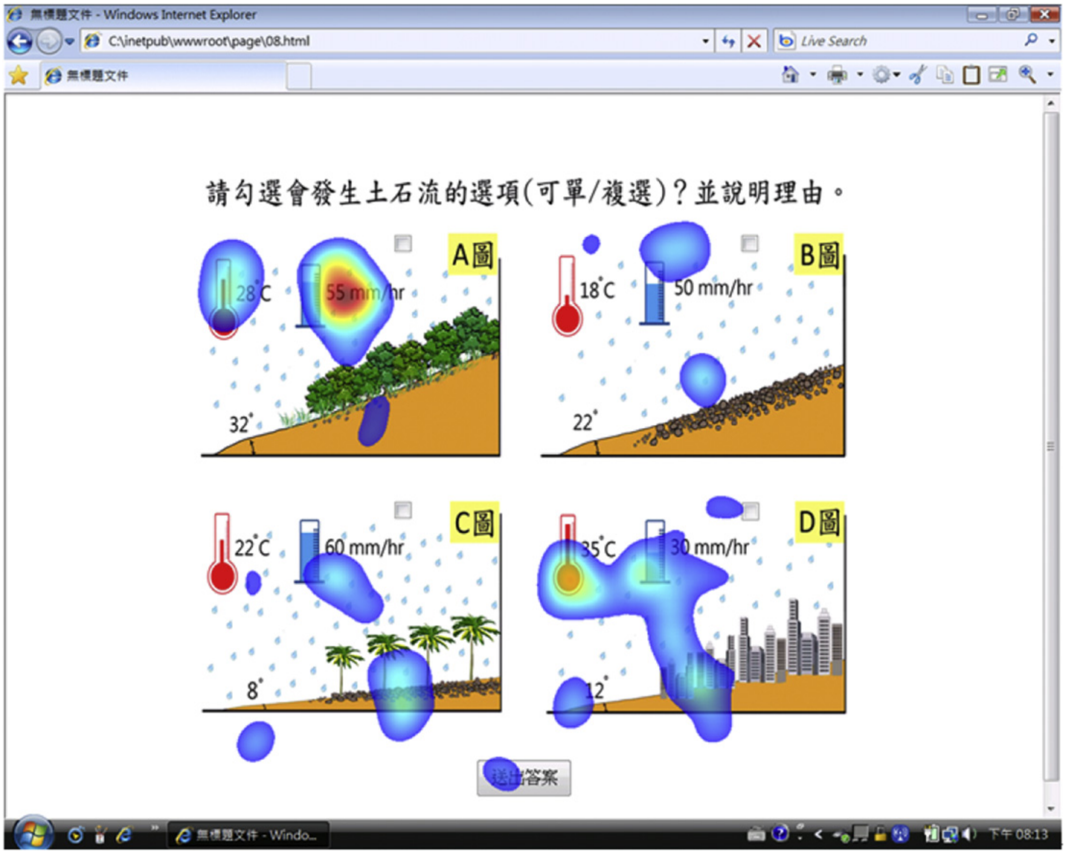

Fig. 6. The Hot Zone image for participant #6 (categorized in low-score performance group, with a response of D) showed that the participant paid most attention on option A and D, especially on factor R and T. The Chinese title states that "*Please select the image(s) inferring a landslide would occur (single/multiple selection) and justify your selection(s).*"

choice problem, students paid most attention to their preferred answers before making decisions. On the other hand, the locations of Hot Zones could provide useful information for a comprehensive understanding of attention distributions in space and might be helpful for predicting the selections made by individuals.

However, an accumulated amount of attention on a particular area seems not enough for understanding individual differences in the process of problem solving. Eye-tracking data can be observed in both a spatial scale and a time scale. A spatial scale represents the areas attended by individuals, and a time scale illustrates the scanning sequences or strategies utilized by individuals for problem solving. Therefore, a sequential analysis on the fixated areas was further conducted to unravel the problem solving strategies used by participants in different performance groups.

4.4. Sequential analysis

In order to further analyze the scan patterns of participants in both performance groups, two *lag sequential analyses* (Bakeman & Quera, 1995; Hou, Sung & Chang, 2009) were conducted for each group on the sequences of LookZones fixated of all participants – one was emphasized on the fixation sequences among *option* LookZones and the other was emphasized on the fixation sequences among *factor* LookZones.

Sequential analysis has been applied to many published studies focusing on behavioral analysis of learning processes (Jeong, 2003; Hou, 2010; Hou, Chang, & Sung, 2009). The analysis method can effectively infer the overall behavioral path patterns during a certain time period of all users. This method was calculated through rigorous conditional probability, expected values matrix, and statistically testing sequential relations between each behavior. Therefore, this study conducted sequential analyses of a group of participants' massive eye movement transition-frequency data during the period of problem solving task, and could infer the characteristics of a group of students' eye

Table 3
Adjusted residuals (z scores) of attentional transitions among *option* LookZones for high- and low-performance group.

	High-sore group						Low-sore group					
	A	B	C	D	Out	Title	A	B	C	D	Out	Title
A	**15.29***	−2.79	−2.18	−5.09	−5.75	N/A	**21.15***	−4.84	−4.5	−5.14	−7.86	−0.95
B	−0.61	**10.77***	−3.19	−1.99	−3.99	N/A	−4.15	**18.45***	−3.58	−3.51	−6	−0.8
C	−4.03	−3.55	**13.49***	−4.08	−1.17	N/A	−4.75	−4.13	**16.5***	−2.18	−3.06	−0.71
D	−5.7	−1.91	−4.46	**15.02***	−3.24	N/A	−5.28	−2.11	−3.24	**14.88***	−2.12	−0.7
Out	−5.43	−1.57	−3.06	−4	**13.39***	N/A	−8.18	−6	−3	−1.8	**16.05***	0.1
Title	N/A	N/A	N/A	N/A	N/A	N/A	−0.95	−0.8	−0.71	−0.7	0.1	**19.01***

*$p < .05$; A, B, C, D, Title = LookZone for option A, option B, option C, option D and problem title, respectively; Out = All other areas besides A, B, C, D and Title.

Table 4
Adjusted residuals (z scores) of attentional transitions among *factor* LookZones for high- and low-performance group.

	High-sore group						Low-sore group					
	T	R	S	D	Out	Title	T	R	S	D	Out	Title
T	**4.78***	**3.55***	−1.02	−1.45	−3.26	N/A	**7.66***	1.27	0.17	−0.83	−3.96	−0.55
R	0.59	**10.7***	−1.45	−2.06	−5.32	N/A	**2.39***	**7.84***	−0.7	−1.39	−5.19	−0.84
S	0.04	−1.43	**9.43***	0.35	−3.5	N/A	−0.97	−1.46	**2.04***	−1.22	0.97	**1.96***
D	−1.45	−0.31	1.19	**3.61***	−1.83	N/A	−1.57	0.13	−1.22	**3.2***	−0.6	−0.7
Out	−1.82	−7.86	−3.47	−0.31	**7.7***	N/A	−3.81	−5.29	0.44	0.08	**5.85***	−3.05
Title	N/A	N/A	N/A	N/A	N/A	N/A	−0.55	−0.83	−0.43	−0.7	−2.15	**16.21***

*$p < .05$; T, R, S, D, Title = LookZone for factor T, factor R, factor S, factor D and problem title, respectively; Out = All other areas besides T, R, S, D and Title.

movement transition paths. As shown in Tables 3 and 4, for option LookZones and for factor LookZones respectively, we derived the Adjusted Residuals Table (Z score table) of high-score group's and low-score group's attentional transition matrices from calculating a series of matrices (including frequency transition matrices, conditional probability matrices, and transition expected-value matrices) (Bakeman & Quera, 1995; Hou, 2010). The rows in the tables represent the *entry LookZones* (i.e. the starting LookZones) and the columns represent the *follow-up LookZones* (i.e. the LookZones that follow the *entry LookZones* subsequently by the learners). Some data are shown as not available here in the tables was because that none of fixations were recorded in the related LookZones. A significant z score in a cell indicates a significant attentional transition from an entry LookZone (labled in row) to a follow-up LookZone (labled in column).

Based on examining Z score of the frequency of each LookZone immediately following another LookZone in Tables 3 and 4, all sequences that have reached a statistically significant level ($p < .05$) for both groups were included in Fig. 7 (among fixated options) and Fig. 8 (among fixated factors). The arrow represented the direction of the transfer, and these four diagrams allowed us to explore the two groups' scan patterns in the problem solving process.

According to Fig. 7, it was clear that no significant difference was found in the inspecting patterns among options between students with high and low performances, except for the difference in inspecting the title. Specifically, participants in the low-score group read the title repeatedly which may indicate that they have difficulties in comprehending the problem. Except that, students in both groups just inspected carefully within each option and no significant sequence was found between any two options in any group. That is, no significant different scan pattern among options was found between successful and unsuccessful problem solvers. Therefore, the Hypothesis 3 is rejected.

However, when comparing the diagrams shown in Fig. 8, two differences were shown in the scanning sequences among factors between the two groups of students. First, high-score students did not pay much attention to the problem title, but low-score students did, especially after inspecting the Slope factor. Secondly, high-score students tended to shift their attentions from Temperature to Rainfall (T–>R, irrelevant to relevant factor, $p < .05$); on the contrary, low-score students tended to shift attentions from Rainfall to Temperature (R–>T, relevant to irrelevant factor, $p < .05$) and (S–>Title, irrelevant factor to problem statement, $p < .05$). These may imply that high-score students focused more on relevant factors of the problem, while low-score students may not only have difficulties in comprehending the problem but also have difficulties in distinguishing and concentrating on relevant factors when solving the problem. This result supported for the Hypothesis 4 that successful and unsuccessful problem solvers inspected the factors in a different manner.

5. Discussion and conclusion

The findings of this study are twofold. First, students, in general, pay more attention to chosen options than to rejected alternatives, and tend to spend more time inspecting relevant factors than irrelevant one, which is consistent with Hegarty et al. (1992)'s finding that relevant information was fixated longer within a text-based problem solving context. Future study can further investigate how students solve science problems with both images and texts by well-designed eye-tracking experiments. As for the longer fixation located in participants' chosen options than rejected options, future research can be done to further explore the reliability and usability of diagnosing or predicting students' question responses based on their visual attention allocations in a larger scale.

Second, there is a significant difference in students' scan sequences among *factors* for solving the problems between successful and unsuccessful problem solvers. Successful problem solvers tend to shift their visual attention from *irrelevant* factors to *relevant* factors, while

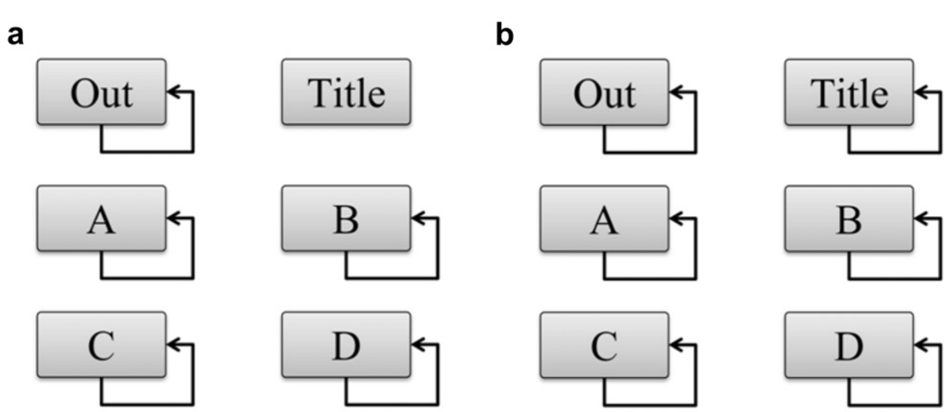

Fig. 7. Sequential analyses among the fixated options for (a) the high-score and (b) the low-score performance groups. No significant difference was found in the attention transformation among options between the two groups.

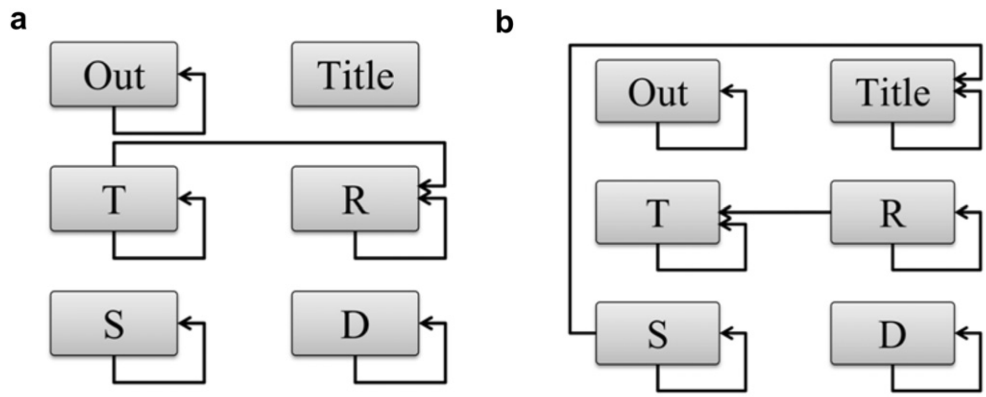

Fig. 8. Sequential analyses among the fixated factors for (a) the high-score and (b) the low-score performance groups. High-score students tended to shift attention from T–>R (irrelevant to relevant) while low-score students tended to shit attention from R–>T (relevant to irrelevant) and S–>Title (relevant to problem statement).

unsuccessful problem solvers tend to shift their visual attention from *relevant* to *irrelevant* factors and to the *problem statement*. It is obvious that the two groups shift attentions in *opposite* directions. This provides a direct evidence to exhibit students' *metacognitive* strategies of different levels profiled in the Model of Strategic e-Learning (Tsai, 2009). The successful problem solvers, with higher levels of metacognitive strategies, are able to *recognize* and *concentrate* on *relevant cues* in a problem-solving learning task. The unsuccessful problem solvers, with lower levels of metacognitive strategies, have difficulties in *comprehending* the goal of a task, *distinguishing* relevant factors from irrelevant factors, and *concentrating* on handling the relevant factors to solve a problem. Therefore, this study concludes that, during an image-based problem-solving learning task, successful learners have higher levels of e-learning strategies than unsuccessful learners. Future studies can use eye-tracking techniques to deeply examine students' metacognitive strategies or cognitive process in different e-learning contexts.

In addition, one finding for unsuccessful problem solvers in this study is similar to Tai et al.'s (2006) finding that some students need more fixations to process information in the zone of problem statements or title descriptions. Given that students' prior knowledge and expertise had influences on allocating visual attention on relevant information in comprehending conceptual graphics (Canham & Hegarty, 2010; Cook et al., 2008), unsuccessful problem solvers might have hold some misconceptions about the required conditions to cause debris slide hazards. Therefore, future study can examine the impacts of students' prior knowledge on their attention allocation and transformation for solving a multiple-choice science problem.

Two major implications can be derived from the findings of this study. First, the characteristics of attention allocations and transformations found in this study for successful and unsuccessful problem solvers provide science teachers some suggestions in instructional practice and curriculum design. The study shows that unsuccessful problem solvers have difficulties in comprehending problem, distinguishing relevant information and concentrating on handling the relevant factors (or ignoring the irrelevant factors) to solve a science problem. This suggests that successful information selection is essential for successful problem solving. Therefore, science teachers should spend more time helping students distinguish relevant information from irrelevant information in addition to explaining the goals of problems. Asking students to explicitly report their information selection and the criteria used for information selection may help students develop their problem solving skills and help teachers realize possible misconceptions.

The other implication of this study concerns future online assessment and instructional system design. Online assessment systems were generally available and often applied in teaching and learning systems (Lazarinis, Green, & Pearson, 2010; Lopez-Cuadrado, Perez, Vadillo, & Gutierrez, 2010; Su & Wang, 2010). Although some of the systems have instant recording and analysis functions (e.g., Su & Wang, 2010) and adaptive testing functions (e.g., Huang, Lin, & Cheng, 2009; Lazarinis et al., 2010), a system with a function to track students' attention is still not available. Future online assessment systems may consider to provide teachers a function to track students' attention allocation and transformation. Provided with this function, teachers can understand more precisely about students' online learning process, for example, the misconceptions that students could have hold as well as students' strength or drawbacks in utilizing metacognitive strategies (e.g., problem solving strategies). And teachers can provide students timely prompts and individual guidance, especially for students who have difficulties to focus on relevant information for learning. Furthermore, the design of automatic diagnosis and feedback mechanisms in online dynamic assessment systems (Moore-Brown, Huerta, Uranga-Hernandez, & Pena, 2006; Wang, 2010) could be further developed according to learners' attention allocations and transformation data.

In sum, this study explored students' visual attention during a complex problem solving task on both spatial and temporal scales. On the spatial scale, the findings of attention allocations help teachers and educational software designers to diagnose students' potential misconceptions or difficulties in problem solving and further design suitable instructional guidance. On the temporal scale, the findings about sequential analyses of fixated zones help researchers and instructors understand students' problem solving strategies such as control and concentration etc. metacognitive strategy. More and more online problem-solving instructional activities/strategies were deeply explored and emphasized (Chen, 2010; Hou, 2011; Laxman, 2010; Tsai, Hsu, & Tsai, in press). Therefore, future studies may apply eye-tracking techniques to explore the cognitive process of online learning and investigate the potential and limitation of its applications in future online assessment and learning systems.

Acknowledgments

This study is supported by National Science Council of Taiwan under the following grants: NSC 98-2511-S-011-004-MY2, 98-2511-S-003-047-MY2, 99-2511-S-011-006-MY3 and 100-2918-I-011-001. The authors appreciate Dr. Hung-Ta Pai's help in MATLAB programming, Dr. Bradley Love's comments and the reviewers' suggestions for revisions.

References

Alkan, S., & Cagiltay, K. (2007). Studying computer game learning experience through eye-tracking. *British Journal of Educational Technology, 38*(3), 538–542.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). UK: Cambridge University Press.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ.* New York: Cambridge University Press.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*, 155–166.

Chen, C.-H. (2010). Promoting college students' knowledge acquisition and ill-structured problem solving: web-based integration and procedure prompts. *Computers & Education, 55*(1), 292–303.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cook, M., Wiebe, N. E., & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphic with macroscopic and molecular representations. *Science Education, 92*(5), 848–867.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455–470.

Evans, M. A., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading: evidence from eye movement monitoring. *Psychological Science, 16*(11), 913–920.

Gandini, D., Lemaire, P., & Dufau, S. (2008). Older and younger adults' strategies in approximate quantification. *Acta Psychologica, 129*(1), 175–189.

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*, 95–99.

Hegarty, M., Mayer, R. E., & Green, C. (1992). Comprehension of arithmetic word problems: evidence from students' eye fixations. *Journal of Educational Psychology, 84*(1), 76–84.

Hegarty, M., Mayer, R., & Monk, C. (1995). Comprehension of arithmetic word problems: a comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*(1), 18–32.

Hou, H. T. (2010). Exploring the behavioral patterns in project-based learning with online discussion: quantitative content analysis and progressive sequential analysis. *Turkish Online Journal of Educational Technology, 9*(3), 52–60.

Hou, H. T. (2011). A case study of online instructional collaborative discussion activities for problem solving using situated scenarios: an examination of content and behavior cluster analysis. *Computers & Education, 56*(3), 712–719.

Hou, H. T., Chang, K. E., & Sung, Y. T. (2009). Using blogs as a professional development tool for teachers: analysis of interaction behavioral patterns. *Interactive Learning Environments, 17*(4), 325–340.

Hou, H. T., Sung, Y. T., & Chang, K. E. (2009). Exploring the behavioral patterns of an online knowledge-sharing discussion activity among teachers with problem-solving strategy. *Teaching and Teacher Education, 25*, 101–108.

Huang, Y.-M., Lin, Y.-T., & Cheng, S.-C. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers & Education, 52*(1), 53–67.

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

Hyönä, J., Lorch, R. F., Jr., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: evidence from eye fixation patterns. *Journal of Educational Psychology, 94*(1), 44–55.

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

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

Jeong, C. A. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussions. *The American Journal of Distance Education, 17*(1), 25–43.

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

Justice, L. M., Skibbe, L., Canning, A., & Lankford, C. (2005). Preschoolers, print and storybooks: an observational study using eye movement analysis. *Journal of Research in Reading, 28*(3), 229–243.

Kaakinen, J. K., Hyona, J., & Keenan, J. M. (2002). Perspective effects on on-line text processing. *Discourse Processes, 33*, 159–173.

Laxman, K. (2010). A conceptual framework mapping the application of information search strategies to well and ill-structured problem solving. *Computers & Education, 55*(2), 513–526.

Lazarinis, F., Green, S., & Pearson, E. (2010). Creating personalized assessments based on learner knowledge and objectives in a hypermedia Web testing application. *Computers & Education, 55*(4), 1732–1743.

Lopez-Cuadrado, J., Perez, T. A., Vadillo, J. A., & Gutierrez, J. (2010). Calibration of an item bank for the assessment of Basque language knowledge. *Computers & Education, 55*(3), 1044–1055.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*, 167–171.

Moore-Brown, B., Huerta, M., Uranga-Hernandez, Y., & Pena, E. D. (2006). Using dynamic assessment to evaluate children with suspected learning disabilities. *Intervention in School and Clinic, 41*, 209–217.

Paulson, E. J., & Jenry, J. (2002). Does the degrees of reading power assessment reflect the reading process? An eye-movement examination. *Journal of Adolescent & Adult Literacy, 46*, 234–244.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: past controversies, current issues, and an agenda for future research. In R. Radach, A. Kennedy, & K. Rayner (Eds.), *Eye movements and information processing during reading* (pp. 3–26). New York, NY: Psychology Press.

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

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension process in reading. *Scientific Studies of Reading, 10*(3), 241–255.

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.

Su, C. Y., & Wang, T. I. (2010). Construction and analysis of educational assessments using knowledge maps with weight appraisal of concepts. *Computers & Education, 55*(3), 1300–1311.

Tai, R. H., Loehr, J. F., & Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessment. *International Journal of Research & Method in Education, 29*(2), 185–208.

Tsai, M.-J. (2009). The model of strategic e-Learning: understanding and evaluating students' e-learning from metacognitive perspectives. *Educational Technology & Society, 12*(1), 34–48.

Tsai, M.-J., Hsu, C.-Y., & Tsai, C.-C. Investigation of high school students' online searching performance: The role of implicit and explicit approaches. *Journal of Science Education and Technology*, in press.

Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare word problems: as eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology, 84*, 85–94.

Wandell, A. B. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.

Wang, T.-H. (2010). Web-based dynamic assessment: taking assessment as teaching and learning strategy for improving students' e-Learning effectiveness. *Computers & Education, 54*(4), 1157–1166.