Scientific visualizations based on integrated model of text and picture comprehension via eye-tracking

Azam Majoonia, Mona Masooda*, Amir Akhavanb

a Centre for Instructional Technology and Multimedia, Universiti Sains Malaysia, Malaysia
b School of Computer Sciences, Universiti Sains Malaysia, Malaysia

Abstract

This study investigates the “color combination effect” and the “sequence of pictures and text effect” in the “Integrated Model of Text and Picture Comprehension” (ITPC) on the formation of the mental model and cognitive load in both desktop and mobile devices. These effects are studied using an experiment including scientific contexts with specific color combinations and different picture and text sequences. The participants of this experiment are graduate students from different fields of science in Universiti Sains Malaysia. Four tasks are assigned for each participant in each phase and every task includes a scientific context accompanied with visualization. After each task the participant is asked to answer a set of questions related to the context. Eye-tracking methods are adopted to record gaze data of the participants while reading each context. Analysis of the gaze data and the percentage of the correct answers to the questions indicate that the sequence of picture and text and use of color in the visualizations results in reduction of the cognitive load of the participants.

Keywords: eye tracking; picture-text sequence; visualization; cognitive load

1. Introduction

The main intention of information visualization is to provide insight about the information in order to solve some prominent problems such as information overload and misinterpretation of data. Several empirical studies have shown the superiority of the visual representations versus the sequences of verbal representations in different tasks such as trend recognition and relation illustration. In the Integrated Model of Text and Picture Comprehension (ITPC), developed by Schnotz, it is claimed that formation of the comprehension of text and visuals are considered as active processes of “coherence formation”, but at the same time it mentions that combination of textual and visual data should be implemented carefully, as it might cause interference in...
the learning of the users because of inappropriate graphics (Schnotz & Kürschner, 2007). In the scientific domains, information visualization is used extensively to provide better representation for the verbal contents (Börner, Chen, & Boyack, 2005; Ware, 2012). The terms, “better representation” is not a precise explanation and there exist a demand for a more explicit comparison parameter to describe the quality and effectiveness of the visualization (Few, n.d. pp.24-26). In order to achieve a general instruction in this study, effect of the parameters, color and the sequence of pictures and text in the formation of the mental model is studied. To provide a comparison between desktop and mobile devices the experiment is conducted on both mediums as nowadays, millions of people use mobile devices for different tasks such as browsing the Internet, reading email, books, blogs and even scientific papers (Maurer, Hausen, De Luca, & Hussmann, 2010). The participants of this experiment are graduate students from different fields of science at Universiti Sains Malaysia. The first phase of the experiment studies the “Picture-text sequencing effect” and “Color combination effect” in scientific contexts containing visualizations on the desktop computer. At the same time the participants are asked to read four short scientific contexts and their gaze data is collected using the eye-tracking devices. Also participant’s comprehension from the context is evaluated using a set of questions at the end of each context. The second phase of the experiment aims to study the “Picture-text sequencing effect” and “Color combination effect” on the mobile devices of small and medium-sized screen. In the second phase, the steps of the first phase are repeated on the mobile devices and likewise the gaze data and the answers to the questions are collected and analyzed. In the final phase the participants are asked to complete the Felder-Silverman learning style model test in order to identify preferred learning style of the participants.

1.1. Related work

In the past decades, a great amount of research has shown that the combination of text and picture provides better learning compared to text only and picture only situations (Anglin, 2004; Fletcher & Tobias, 2005). Several research also is conducted to answer this question that if the text and picture appear in sequence, which model would be more effective. Some of these studies have provided experiments, to study the effect of picture and text sequencing (Canham & Hegarty, 2010; Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013; McCrudden, Magliano, & Schraw, 2011; Shaw, Nihalani, Mayrath, & Robinson, 2012). The results of the conducted experiments varies for each study, meaning that some of the experiments concluded in the effectiveness of picture before text sequence (Borges & Robins, 1980; Dean & Enemoh, 1983; McCrudden et al., 2011; McCrudden, Schraw, & Lehman, 2007; Salmerón, Baccino, Cañas, Madrid, & Fajardo, 2009; Ullrich, 2007), and also almost equally a big body of research have achieved positive results on behalf of the text before picture or text only model (Bransford & Bransford, 1972; Canham & Hegarty, 2010; O’Keefe & Solman, 1987; Robinson, Katayama, Dubois, & Devaney, 1998; Robinson & Kiewra, 1995; Simmons, Griffin, & Kameenui, 1988). The Picture-text sequencing effect mentioned in the ITPC model, claims that the presentation of the picture before text, provides the mind preliminary visual information that helps with the creation of the mental model (Schnotz & Kürschner, 2007).

1.2. Picture and text sequencing effect on desktop and mobile devices

This study is important because, it investigates effect of picture and text sequencing on both the desktop and mobile devices side by side, while examining the influence of color on it. The effect of picture and text sequencing on mobile devices can be of more importance, as the screen size of the mobile devices are generally smaller than the desktop computers; which in this case when the picture is accompanied by text (or vice versa), the content producer is limited in the display size, thus there is no other way but to present picture and text in a sequential format.

2. Experiment

This study aims to determine the effect of color combination and image/text sequencing on the understanding and cognitive load of postgraduate students while reading scientific contexts on desktop and
mobiles devices. In order to investigate the effect of these variables, an experiment is designed with two parts including four tasks to examine the cognitive load of the students using the desktop monitor and a mobile device.

2.1. Participants

Participants of this study are 10 third year Universiti Sains Malaysia postgraduates with a science background aged between 29 to 35 years old. The participants are assured that their test results would be anonymous and their score will not be disclosed. The participants are divided into 2 groups, each group with eight similar texts but different sequence of picture and text.

2.2. Apparatus

In this experiment, the main goal was to investigate the cognitive load, of the participant in the picture first, followed by text and then text first followed by picture, on both desktop and mobile devices. In order to achieve this goal, Gaze data was collected using SMI Red, remote eye-tracking device in the desktop based tasks and SMI eye-tracking glasses (ETG) for the mobile device based tasks. The calibration of the devices are renewed after each task and the participants are asked to answer a set of questions related to the tasks.

After completing the tasks the participants are given the Felder-Silverman learning style questionnaire, to determine if they are visual or verbal learners.

2.3. Materials

Material used in this experiment, are eight simple passages which provides description on the results of the survey conducted on the academic databases. The length of each passage is between 150 to 210 words and the readability score for each of the passages is calculated using the readability formula (Kincaid, Fishburne Jr, Rogers, & Chissom, 1975) and the mean of the readability score is 14.2 with below 10% standard deviation. The material is accompanied with diagrams related to the context.

2.4. Procedure

The participants, initially are informed that they would read four tasks which include a short passage with a picture before or after the passage, on desktop computer and four tasks on mobile device and after each task they will answer a few questions related to each task.

The experiment includes three major sections:

(a) Reading tasks with questions on the desktop with remote eye-tracking device (RED):
Four reading tasks are assigned for each participant and after each task participants are asked to answer a set of questions about the context and their gaze data are recorded using the remote eye-tracking devices.

(b) Reading tasks with questions on the mobile devices with eye-tracking glasses (ETG):
The same procedures in step (a) are implemented, except this time the tasks and the questions are provided using small and medium-size screen devices and their gaze data are recorded using eye-tracking glasses.

(c) Felder-Silverman learning style model test:
In the final step, the participants are asked to answer 44 questions of the “Felder-Silverman learning style model” to identify their Learning Style Index (LSI) for further analysis. Fig. 1 demonstrates the schematic process of the experiment.
3. Results

The collected information from the experiment can be divided into four categories:

- Scores of the questions about the context on the desktop and mobile device
- Eye-tracking results of the experiment on the context on the desktop
- Eye-tracking results of the experiment on the context on the mobile device
- Results of the Learning Style Index test

The analysis of the collected information is completed in three steps, firstly, the learning style of the participants is identified and the attributes of the visual/verbal learner are added to the database of the collected information. In the second phase, the mean score of the questions for all the participants for each test is calculated. Later the mean results for each group of visual and verbal learners are also calculated separately. Then the gaze data are plotted using the heat map and the scan path visualization methods to investigate the fixations and movement sequence of the eyes collected by the eye-tracking devices.

3.1. Gaze data

For further analysis, the scores and gaze data collected from the desktop and mobile devices are clustered and the results are analysed using classification method. The eye-tracking data provides information about the time needed for the participant to read and recognize each part of the context. It can also help estimate the cognitive load of the participant while going through the context. The collected eye tracking data are usually visualized either by heat map and scan path or by analysing the number and duration of the fixation on an Area of Interest (AOI) or POI. However the other eye tracking parameters such as blinking, pupil size variation and length of the saccades, ratio of fixations, are also very important parameters (Bixler & D’Mello, 2014). In this study, the parameters for each task and for each participant are put into a matrix of the eye tracking data, and 50% of all the features are recorded in this matrix are used for training the model using a supervised learning method. In the next step, the rest of the records are used for testing the hypothesis on the effectiveness of picture-text sequencing model.
3.2. Collected data on the experiment

The analysis of the results are completed using a method based on the records of the fixation, saccades, blinks and pupil size on two large AOIs. In order to investigate the relation of the cognitive load and gaze data in a quantitative manner, all the eye tracking data are summarized into 7 features and 2 matrices. This matrix of gaze data contains: (1) Time ratio of the Fixations, (2) Mean of the Fixations, (3) Number of Saccades, (4) Number of blinks, (5) Mean time between blinks, (6) Mean pupil diameter, and (7) Total length of eye movement. The results of the answers to the questions are used as the classifier attribute for the training step, and then this attribute is again used for testing the model. The ZeroR classifier in the WEKA data mining toolkit (Hall et al., 2009) is used for the training and testing. In this process, 50% of the participant’s data (attribute matrices) are selected randomly for training purpose, 10% for validation and 40% for testing.

3.3. Normalization of the data

The data in the matrices for each participant are not all in one range, as each attribute represents one different feature, also physical characteristics (such as pupil diameter) vary for each participant, which leads to a bias in the modeling. In order to avoid this, the data inside the matrices are all normalized using a common method of normalization (in this case Hyperbolic Tangent function) therefore all the parameters will have values in the same range of 0 and 1. Fig. 2 is the matrix of the normalized features for participant one from group one.

| The features matrix for participant one, after normalization |
|----------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Mobile device (ETG)              | Tasks         | Time ratio of Fixations | Mean of the Fixations | Number of Saccades | Mean time between blinks | Mean pupil diameter | Total length of eye movement |
| Desktop computer (RED)           | 1             | 0.386              | 0.731             | 0.709             | 0.484             | 0.158           | 0.169              | 0.353 |
|                                 | 2             | 0.377              | 0.867             | 0.687             | 0.740             | 0.172           | 0.177              | 0.591 |
|                                 | 3             | 0.396              | 0.6415            | 0.822             | 0.596             | 0.370           | 0.595              | 0.176 |
|                                 | 4             | 0.395              | 0.5918            | 0.722             | 0.555             | 0.449           | 0.211              | 0.169 |

Fig 2. The features matrix for participant one after normalization for modelling

3.4. Results of the questions

The participants are asked to answer questions related to the context in each task, and the results of these questions are used as a measure of the understanding of the participants from the context. The mean of the results for all the participants is presented in Table 1 and Table 2. Although the number of participants was not too many, the results provide clues about the effectiveness of the use of color. Also, better results are achieved in the Picture first cases compared to the Text first cases, which supports the ITPC theory.

| Table 1. Results of the questions on Mobile device (Mean results of Groups 1 and 2) |
|----------------------------------|----------------|----------------|----------------|----------------|
| Context                          | Group 1 Colored | Group 1 Black and White | Group 2 Colored | Group 2 Black and White |
| Picture first                    | 90%            | 80%            | 80%            | 80%            |
| Text First                       | 65%            | 70%            | 65%            | 60%            |
Table 2. Results of the questions on Desktop computer (Mean results of Groups 1 and 2)

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Colored Black and White</td>
<td>Colored Black and White</td>
</tr>
<tr>
<td>Picture first</td>
<td>90%</td>
<td>85%</td>
</tr>
<tr>
<td>Text First</td>
<td>67%</td>
<td>75%</td>
</tr>
</tbody>
</table>

3.5. Analysis of the pupillary response

According to Klingner, Kumar, and Hanrahan (2008), pupil diameter increases with the difficulty of the context. In this experiment, the pupil data for each of the participants, for each task on two main AOIs are collected and analyzed, the Fig 3 demonstrates changes of the pupil size while the participant is reading the context.

Fig 3. (a) the pupil diameter increases while reading the text and (b) during the process of reading pupil diameter is higher than while gazing at the figure

3.6. Analysis and classification

A parameter, representing the cognitive load is presented in this section which represents the cognitive load score (CLS). This parameter, is achieved by normalizing the maximum score of the parameters monitoring the cognitive load (number of blinking, size of pupil and fixation). In other words, values of the eye tracking features while watching a black space and while trying to read and understand a difficult section of the tasks is scored (between 0% and 100%). The mean value of the all features are tested using the model
and the similarity factor is considered as the mean cognitive load for the tested tasks. Table 3 shows the results of the analysis.

The results of the cognitive load score, reflects the difficulty of the contexts for all the participants (Table 3), the scores suggests that, sequence of picture before text, can reduce the mean of the cognitive load. Also the mean of CLS in the contexts with colored figures is noticeably lower than the contexts with black and white figures.

Table 3. Mean cognitive load score from testing of the model trained by 50% of the participants

<table>
<thead>
<tr>
<th>Device</th>
<th>Tasks</th>
<th>Cognitive load score</th>
<th>Tasks</th>
<th>Cognitive load score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile device</td>
<td>Picture first, Colored</td>
<td>69%</td>
<td>Picture first, B/W</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>Text first, Colored</td>
<td>74%</td>
<td>Text first, B/W</td>
<td>81%</td>
</tr>
<tr>
<td>Desktop computer</td>
<td>Picture first, Colored</td>
<td>59%</td>
<td>Picture first, B/W</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>Text first, Colored</td>
<td>70%</td>
<td>Text first, B/W</td>
<td>78%</td>
</tr>
</tbody>
</table>

4. Discussion

In this work, the effect of the color and the sequence of picture and text in the formation of the mental model and cognitive load is investigated for both the desktop and mobile device. The results support the Integrated Model of Text and Picture Comprehension on the effect of the sequence of the picture and text in both mobile devices and desktop computers. Also the results confirm the positive effect of color in the reduction of cognitive load in all the cases. Results of the Felder-Silverman test, provided that 75% of the participants were visual learners and 25% verbal learners. Nevertheless, in this experiment both verbal and visual participants had better scores on the picture first tasks.

5. Conclusion

The effect of sequence of pictures and text in the cognition load and the formation of the mental model in the scientific contexts in the Integrated Model of Text and Picture Comprehension is a rarely investigated issue and in this experiment, the authors try to conduct an analysis regarding this effect on both the desktop and mobile devices. The contribution of this study is both on the investigation of the theory and also technically analyzing the effect of using mobile devices versus desktop computers.

Acknowledgements

This research is supported by the Center for Instructional Technology and Multimedia and the Research University Grant awarded by Universiti Sains Malaysia.

References


