Exploring Students’ Eye Movements to Assess Learning Performance in a Serious Game

Asma Ben Khedher, Imène Jraidi and Claude Frasson
University of Montreal
Montreal, Quebec, Canada
{benkheda, jraidiim and frasson}@iro.umontreal.ca

Abstract: In this paper, we propose to analyze learners’ gaze data as they interact with a learning environment. The primary goal of this research is to assess students’ learning experience and to investigate what eye movements can tell about learners’ performance. We collected data from 15 medicine students when resolving clinical tasks in a medical education game. Results revealed that students’ visual behavior differs from one learner to another depending on the medical case solved. However, gaze data showed that there is a particular area of interest that raised the greatest attention. Interestingly, we found that the fixation duration spent on this area has the least positive effect on the learners’ performance.

Keywords: Eye Movements, Students’ Performance Assessment, Learning environment

Introduction

Computer-based environments consistently seek to assess users’ behavior and monitor their interaction experience. In this context, the use of non-intrusive physiological sensors has gain much interest in the last decade as they provide reliable and unbiased information (Chaouachi, Chalfoun, Jraidi, & Frasson, 2010; Jraidi, Chaouachi, & Frasson, 2013; Jraidi & Frasson, 2013). In particular, using eye tracking to assess users’ experience has shown considerable promising results. Researchers seek particularly to investigate what gaze data can reveal about users’ visual patterns when interacting with the system. In fact, web design and web pages’ usability are among the research domains where eye tracking is the most applied as a mean to understand users’ activities throughout web search in an effort to provide a more ergonomic web interface and facilitate thus browsing (Cutrell & Guan, 2007; Jankowski et al., 2016). In the same context, eye tracking is used in human-computer interaction to study usability issues (Poole & Ball, 2005; Weill-Tessier, Turner, & Gellersen, 2016) and in visualization tasks to provide effective user-adaptive systems (Jaques, Conati, Harley, & Azevedo, 2014; Lallé, Conati, & Carenini, 2017). The development of computer-based adaptive systems using eye tracking relies principally on the assessment of users’ visual behavior, users’ visual pattern and users’ experience during the interaction (Mehigan et al., 2011; Raptis et al., 2017) and that is what learning environments seek to do consistently to enhance students’ learning progress.

In this paper, we aim to examine the feasibility of using eye tracking as a tool to analyze learners’ visual behavior to investigate what eye movement fixation-based metrics can tell us about students’ performance. This experimental study is part of a larger research that explores the feasibility of eye tracking technology in assessing students’ visual patterns where participants’ eye movements were recorded during problem-solving tasks.

The rest of the paper is structured as follows: section 2 presents some existing works, section 3 presents our learning environment and the experimental design. In section 4, we present the obtained results and finally, in section 5, we conclude with some future works.

Related Work

Computer-based learning environments use different tutoring strategies in order to enhance the students’ experience and optimize their learning outcomes (Ghali, Frasson, & Ouellet, 2016; Jraidi, Chalfoun, & Frasson, 2012; Jraidi & Frasson, 2010). Assessing the students’ visual behavior during the learning process has a potential impact on the effectiveness of these strategies. Indeed, eye tracking systems provide valuable insight into the learners’ visual patterns during cognitive task resolution. This can be used either in the design of the learning interface whereby students
interact, or during the learning process by providing adequate support when needed. In (Pachman et al., 2016), the authors used eye tracking to detect learners’ confusion in digital learning environments. They found that confusion ratings were positively correlated with the fixations on the non-relevant areas, which helped them detect potential confusion areas with the future goal of setting self-regulatory techniques for students to self-manage their confusion during learning.

Raptis and his colleagues investigated students’ visual activities in order to infer cognitive differences among individuals when performing different visual tasks. They were able to discriminate between field-independent and field-dependent learners. In visual search tasks, the first class of users performed an organized and well-oriented path when searching. For the visual decision-making task, less and shorter fixations were found for the field-dependent learners. These results would be useful to adapt learning environments to learners’ cognitive styles (Raptis et al., 2017).

Using eye tracking in serious game studies is also gaining a considerable success over the past years (Ben Khedher, Jraidi, & Frasson, 2017a, 2017b, 2018). The aim of educational game communities is to offer a pleasant and effective learning environment for the students in order to keep them immersed (Kickmeier-Rust, Hillemann, & Albert, 2011; Kiili, Ketamo, & Kickmeier-rust, 2014). In the work of Byun et al., the authors analyzed learners’ in-situ data to investigate how eye tracking can be used to evaluate students’ performance while they are playing a serious game. Scan path and area of interest (AOI) based metrics were used in the study to differentiate between novice and expert players. Results showed that experts have a more organized visual path and tend to focus more on the relevant areas in the game compared to the novices who have a randomly visual trajectory (Byun, Loh, & Zhou, 2014).

In this paper, we propose to use eye tracking as a tool to track and assess students’ visual experience in order to evaluate their learning outcomes. In fact, eye movement metrics such as fixations and saccades influence the learners’ progress and cognitive performance and consequent student learning. However, eye movements and learners’ performance may be negatively correlated. Longer fixations, for example, are not always indicators of learning success and greater attraction.

**Experimental Design**

A study was conducted to track students’ eye movements while they were interacting with our medical learning environment called Amnesia. We aim to identify, first, whether there were particular elements that have attracted the students’ interest more than others. Second, we seek to investigate whether these elements potentially contribute to the students’ success or failure when resolving the medical cases.

**Amnesia**

Amnesia is a medical serious game developed for novice undergraduate medicine students to assess their clinical abilities through problem-solving tasks. The environment features a virtual hospital where the player is a doctor who was by mistake diagnosed with amnesia and taken into a trap within the hospital. Participants need to prove in a first stage, that they do not suffer from amnesia by resolving some cognitive tasks such as logic tests (e.g. number series and analogies) and in a second stage, they have to prove their clinical abilities by resolving six medical cases. These tasks were examined and validated by a medical professional.

For each case, the main objective is to identify the correct diagnosis and treatment by considering a series of observations. The different diseases to be identified are flu, bacterial pneumonia, measles, ebola, mumps and whooping cough. The player can establish a first diagnosis based on the demographic information and the symptoms exposed by the patient. The students can make an initial hypothesis and they have up to three attempts to find out the right response. If the initial diagnosis established is not valid, the players can collect additional clinical data such as analyses and antecedents until reaching the correct diagnosis. Once they found the appropriate one, the students are also given three attempts to identify the adequate treatment and after three errors made either in the diagnosis or the treatment, the game is over.

**Participants and Apparatus**

Fifteen undergraduate medicine students (8 males) aged between 20 and 27 (M = 21.8 ± 2.73) participated voluntarily in the experiment. A Tobii Tx300 eye tracker was used to record participants’ eye movements using a
sampling rate of 300 Hz. It integrates infrared sensors and a camera to record the session on a 23-inch computer monitor with a (1920 x 1080) resolution. The Amnesia environment was presented on this eye tracking system which was approximately placed at a distance of 65 cm from the participant. A nine-point calibration grid was used to evaluate eye gaze quality.

In the present study, we used areas of interest (AOIs) to obtain a more detailed analysis. Tobii permits to manually create AOIs that represent a specific screen region and to report gaze data relative to each area. We defined six areas representing relevant regions of the screen denoted as follows: Information (I), Antecedents (A), Symptoms (S) and Analysis (N). All the medical cases have the same AOIs. From the obtained data, two fixation-based metrics were computed within each area: fixation duration (F.D) and time to first fixation (T.T.F.F). Fixation duration is the average time spent in an AOI and the time to first fixation indicates the amount of time it takes a respondent to look at a specific AOI from stimulus onset.

Protocol

The experiment was set up in our research laboratory. First of all, students were instructed to sign a consent form explaining the study and the material. Then, they were invited to take place in front of the eye tracker and were informed that free head movements were allowed. Following the calibration process, the experiment starts with a fast introductory scene where the players are briefed about the main objectives of the learning environment and the tasks they need to fulfil within the game. The study lasted about 1 hour during which the participants interacted with the environment during 30-45 minutes and then, they were asked to fill in a post-game questionnaire about the game design and usability in order to collect their feedbacks for further improvements in the game.

Results and Discussion

In this study, our main objective is to track students’ eye movements during the resolution of the medical case to have a general overview of their visual behavior across the different areas of interest in each medical case. We aim to identify which regions attracted their attention the most in terms of fixation-based metrics and whether there is a relationship between the most fixated areas and successful responses. The question is: can we conclude that looking to a particular area for a certain period of time means that this area really caught our attention, or that our gaze has just landed on it without any particular focus?

Since the medical cases were different in term of content, we were interested in exploring the visual behavior of the students across each case separately in order to obtain more reliable findings. We first generated students’ heat maps to have a global visualization of the learners’ eye movements and then we performed statistical comparisons to support our work.

Heat Maps

The eye tracking heat map visualization tool indicates the areas with high density of fixations (Bojko, 2009). Using heat maps, we can identify where the participants focus more their attention within the screen by representing the most frequently visited areas with a red colour, followed by yellow and then green for the areas with less fixation intensity.

Fig. 1 presents six heat maps from the same participant, each associated with a different medical case. We can easily see that the symptoms are the most fixated by the participant, in all the cases, with a fixation duration (F.D) in seconds equal to 18.41. Then, the analysis area comes second, with less attention (F.D = 4.95) followed by the information (F.D = 3.13) and finally the area of antecedents that have shown the lowest attention (F.D = 1.45). Indeed, as illustrated in the participant’s heat maps, the “A” area of interest located on the lower left corner shows little fixation intensity in almost all the cases. However, when taking the case four into consideration, the student exhibited the opposite. He fixated the antecedents’ area (F.D = 4.1 sec) as much as the symptoms’ (F.D = 4.2 sec) and the analysis’ areas (F.D = 4.7 sec) for two reasons. First, this is the only medical case where the patient has antecedents. Second, the disease that the student should identify in this case four is Ebola, and among the elements which the student needs to pay attention in order to obtain a good diagnostic, is what is listed in the antecedents’ area saying that the patient has recently travelled to Guinea.
Although in this example the antecedents’ area caught the attention of the learner due to its importance, the symptoms remain the most interesting area in terms of fixation metrics for all participants within all cases. Yet several questions arise. Does looking into an area of interest mean that this area is relevant? Does fixation duration have an impact on students’ performance when solving cognitive tasks? Does looking to a particular area can potentially lead to successful responses?

![Attention heat maps from the same participant in each medical case](image)

**Figure 1.** Attention heat maps from the same participant in each medical case

**Students’ Eye Movements across the AOIs**

Tab. 1 shows the statistical comparison between the different areas using the mean values (\(M\)) and standard deviation (\(SD\)) for the fixation duration metric. The results, we obtained, are also in line with the attention heat maps presented in the previous section. We can clearly identify that the “S” (Symptom) area was by far the most fixated zone by all the participants in all cases. These results are obvious since identifying the patients’ symptoms is among the most important steps in a medical diagnosis process. In contrast, the fixation duration in the “A” (Antecedents) area was shorter than the other areas because in the majority of the cases, patients did not have particular antecedents; this is why this area did not need much attention.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th></th>
<th>A</th>
<th></th>
<th>S</th>
<th></th>
<th>N</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>4.35</td>
<td>2.65</td>
<td>1.79</td>
<td>1.35</td>
<td><strong>16.78</strong></td>
<td>14.64</td>
<td>10.37</td>
<td>7.05</td>
</tr>
<tr>
<td>Case 2</td>
<td>2.85</td>
<td>1.75</td>
<td>0.69</td>
<td>0.61</td>
<td><strong>25.89</strong></td>
<td>12.56</td>
<td>7.25</td>
<td>4.82</td>
</tr>
<tr>
<td>Case 3</td>
<td>2.56</td>
<td>1.79</td>
<td>0.71</td>
<td>0.69</td>
<td><strong>18.69</strong></td>
<td>7.3</td>
<td>4.07</td>
<td>2.17</td>
</tr>
<tr>
<td>Case 4</td>
<td>1.22</td>
<td>1.01</td>
<td>1.68</td>
<td>1.18</td>
<td><strong>13.90</strong></td>
<td>11.48</td>
<td>4.27</td>
<td>1.46</td>
</tr>
<tr>
<td>Case 5</td>
<td>1.60</td>
<td>1.29</td>
<td>0.70</td>
<td>0.69</td>
<td><strong>20.72</strong></td>
<td>10.48</td>
<td>2.93</td>
<td>1.87</td>
</tr>
<tr>
<td>Case 6</td>
<td>1.26</td>
<td>1.39</td>
<td>0.78</td>
<td>0.69</td>
<td><strong>28.72</strong></td>
<td>17.48</td>
<td>5.92</td>
<td>3.07</td>
</tr>
</tbody>
</table>

**Table 1.** Statistical comparison between the AOIs in each medical case in terms of fixation duration.
In accordance with these findings, one-way analyses of variance (ANOVA) were conducted to investigate whether there were significant differences among all the AOIs and in terms of fixation duration and time to first fixation metrics (T.T.F.F). The results showed statistically significant results ($p < 0.001$) among all cases in terms of fixation duration as shown in Tab. 2. However, the T.T.F.F metric showed significant differences only in case one ($p < .001$). We performed Tukey post-hoc tests to compare all areas of interest by pairs in order to identify which area of interest caught the students’ attention the most. The results indicated that the symptoms’ region differed significantly ($p < 0.05$) from the other areas in almost all the cases. Based on all fixation metrics, the time dedicated to looking to the symptoms’ area by the participants far exceeds ($p < .001$) the time dedicated to the other areas proving that the symptoms’ area is a main zone that attracts more attention than the others.

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
<th>Case 5</th>
<th>Case 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F (3, 11)=9.83$</td>
<td>$F (3, 6)=28.52$</td>
<td>$F (3, 6)=43.95$</td>
<td>$F (3, 3)=7.15$</td>
<td>$F (3, 3)=15.73$</td>
<td>$F (3, 1)=15.89$</td>
</tr>
<tr>
<td><strong>Information</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.35 (2.65)</td>
<td>2.85 (1.75)</td>
<td>2.56 (1.79)</td>
<td>1.22 (1.01)</td>
<td>1.60 (1.29)</td>
<td>1.26 (1.39)</td>
</tr>
<tr>
<td><strong>Antecedents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.79 (1.35)</td>
<td>0.69 (0.61)</td>
<td>0.71 (0.69)</td>
<td>1.68 (1.18)</td>
<td>0.70 (0.69)</td>
<td>0.78 (0.72)</td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.78 (14.64)</td>
<td>25.89 (12.56)</td>
<td>18.69 (7.30)</td>
<td>13.90 (11.48)</td>
<td>20.72 (10.48)</td>
<td>23.72 (11.87)</td>
</tr>
<tr>
<td><strong>Analyses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.37 (7.05)</td>
<td>7.25 (4.82)</td>
<td>6.50 (8.15)</td>
<td>4.27 (1.46)</td>
<td>2.92 (1.87)</td>
<td>5.69 (3.07)</td>
</tr>
</tbody>
</table>

Table 2. Descriptive statistic results comparing AOIs in terms of fixation duration in all cases. ANOVAs reported in italics ($p < 0.001$)

**Differences in Visual Behavior Based on Students’ Performance**

In order to foster students’ learning outcomes, we need to identify what are the elements that can possibly lead to successful learning performance. In other words, is there any area(s) of interest that can affect the failure or success of the participants? Recalling that in the previous analysis, the symptoms’ area was the most fixated by all participants in all cases, we then assume that there is a relationship between this area and the performance of the students when solving the medical cases.

In this study, several multivariate analyses of variance (MANOVA) were performed to identify in each case, whether looking to the “S” AOI can help students’ obtain successful responses and if not, which area can then better did. The dependent variables were fixation duration and time to first fixation metrics and the independent variable was success or failure in the diagnosis step.

**Symptoms’ Area.** As shown in Tab. 3, only one significant effect was found in the last case in terms of fixation duration ($p < 0.005$) showing that focusing on the “S” AOI may potentially have an impact on the success or failure. For the first case, we could not conduct the MANOVAs since the number of subjects was not fair between the two groups (group 1: success; group 2: failure). For the remaining cases, the results indicated no statistically significant relationship between the fixation metrics on the symptoms and the identification of the correct diagnosis. However, as Graesser and his colleagues stated in (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005), ineffective learners fixated the relevant areas randomly, therefore, we cannot conclude that longer fixations on the symptoms area do contribute in identifying the right diagnosis.

At first glance, one might think that these results are contradictory, as we found previously that participants pay more attention to the symptoms when reasoning, nevertheless these findings can make sense. One of the possible explanation is that longer fixation duration is not always an indicator of greater interest (Djamasbi, 2014). It can also indicate distraction and confusion regarding the difficulty of the task especially when the participants cannot succeed in finding the right diagnosis. In fact, we must also remember that the participants are novices who are undergraduate medicine students that have not yet acquired all the necessary knowledge. There may be also another possible reason which is the off-task behavior (Rowe, McQuiggan, Robison, & Lester, 2009). This phenomenon is very usual in computerized learning environments where students disengage from the learning content. This disengagement is
manifested either by gaming the system to advance quickly in the game without making any cognitive effort, or simply when the participant become misguided because the problem was difficult. That is, students are just looking into the symptoms without any particular attention since they did not master the content.

<table>
<thead>
<tr>
<th>Case</th>
<th>F.D</th>
<th>T.T.F.F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>F (3, 11)</td>
<td>( p = .14 )</td>
</tr>
<tr>
<td>Case 2</td>
<td>F (3, 6)</td>
<td>( p = .45 )</td>
</tr>
<tr>
<td>Case 3</td>
<td>F (3, 6)</td>
<td>( p = .45 )</td>
</tr>
<tr>
<td>Case 4</td>
<td>F (3, 3)</td>
<td>( p = .77 )</td>
</tr>
<tr>
<td>Case 5</td>
<td>F (3, 3)</td>
<td>( p = .002 )</td>
</tr>
<tr>
<td>Case 6</td>
<td>F (3, 1)</td>
<td>( p = 121.08 )</td>
</tr>
</tbody>
</table>

Table 3. MANOVAs results between the symptoms’ fixation metrics (F.D: fixation duration and T.T.F.F: time to first fixation) and diagnosis performance (success or failure)

**Remaining Areas.** In the previous section, results revealed that the symptoms’ AOI has no effect on the students’ performance when resolving the medical cases. This is why we run additional MANOVAs case by case to show whether such areas exist or not.

In case one, all the participants succeeded in identifying the diagnosis so we discarded the data from the analysis. Case 2 \( F(1, 8) = 4.946, p = n.s \) and 5 \( F(1, 3) = 0.084, p = n.s \) showed no differences between the areas of interest \( (p = n.s) \). For the remaining three cases the results were statistically significant in terms of time to first fixation. Post hoc tests were carried out for separate correlational analyses to demonstrate which AOI could potentially contribute to students’ success. The findings report that there is no unique AOI that leads to learners’ performance, there were different regions of interest in each case separately. For case 3 \( p = 0.001 \) and 6 \( p = 0.031 \) a significant difference was found for all areas: information, analysis and antecedents. In case 4, a significant contribution was found only for the antecedents’ area \( (p = 0.000) \) showing that the obtained results are actually consistent with the medical case design. In fact, in case 4, the antecedents’ area of interest is the most important area since the correct diagnosis is Ebola so, students need to focus particularly on this area which includes relevant information ("recently travelled to Guinea") for the students that may help them find out the right diagnosis.

In summary, eye tracking demonstrated its feasibility in tracking learners’ experience. We were able to identify the areas that caught the most the participants’ attention using fixation metrics and heat maps. The results showed that students’ visual behavior may differ from one person to another depending on the medical case. However, they experience the same focus on the symptoms’ area which attracted higher attention providing, thus a better understanding of how novice medicine students are reasoning within the learning environment. Moreover, we linked these results with the assessment of the students’ performance in order to investigate whether there is a relationship between fixation-based metrics and students’ progress during medical cases resolution. Despite, the students seemed to spend more time fixating the symptoms in all cases, the case-by-case analyses we performed, revealed that the symptoms’ AOI has no impact on identifying the correct diagnosis and that there were other areas that contributed to the learners’ success. These findings provide encouraging evidence on using eye tracking in learning environments as a tool for students’ performance assessment.

**Conclusion**

In this paper, we used eye movements to track students’ visual behavior in order to assess their learning progress. An experimental protocol was conducted while 15 novice medicine students were interacting with the Amnesia medical environment and resolving different medical cases. The objective was to identify the regions that attracted the most their attention. ANOVAs’ results and attention heat maps showed that almost all participants have a similar visual pattern, showing that, the symptoms’ area caught the most interest. Further analyses were performed to investigate whether this area has an impact on students’ performance. Yet, contradictory results have been found
showing that, among other metrics; fixation duration is not always an indicator of higher cognitive load. It can be negatively correlated with students’ success. This experimental study is part of a larger research that has potential implications for learning environments that aim particularly to assess students’ reasoning within problem-solving tasks. Our long-term research objective is to provide students with supportive interventions while reasoning by considering further physiological variables such as EEG engagement and workload.

Acknowledgment

We acknowledge SSHRC (Social and Human Research Council) through the LEADS project and NSERC (National Science and Engineering Research Council) for funding this research.

References


