INTERNATIONAL CONFERENCE ON HUMAN BEHAVIOR IN DESIGN
14-17 OCTOBER 2014, ASCONA, SWITZERLAND

DESIGNING ONLINE TESTS FOR A VIRTUAL LEARNING ENVIRONMENT – EVALUATION OF VISUAL BEHAVIOUR BETWEEN TASKS

Gergely RAKOCZI[1], Andrew DUCHOWSKI[2] and Margit POHL[1]
[1]Vienna University of Technology, Institute for Design and Assessment of Technology, Vienna, Austria, [gergely.rakoczi | margit.pohl]@tuwien.ac.at
[2]Clemson University, School of Computing, Clemson, SC, USA, duchowski@clemson.edu

ABSTRACT
Visual attention of learners is examined during completion of an online e-learning test consisting of three different educational tasks, namely matching, multiple choice, and true/false tasks. Eye movements of 36 learners were recorded with respect to four layout areas of the virtual learning environment's interface, including textual instruction, visual content, selection items, and navigational elements. Visual behavior was analyzed along five traditional eye tracking metrics: total gaze duration, time to first fixation, mean fixation duration, fixation count, and regression rate. Additionally, transition matrices were used to analyze attentional switching between the four layout areas. Comparison of transition matrices facilitates effects analysis between the three different tasks, quantifying significant differences in aggregate visual behavior. Mean fixation durations indicated more visual effort for true/false tasks. The matching task transition matrix suggests a significantly more focused pattern of attentional switching, compared to the other tasks. Based on these findings, a few practical recommendations for designers of e-learning environments are offered.

Keywords: Eye movement, eye tracking, visual behavior, task design, e-learning, human-computer-interaction, virtual learning environments (VLE)

1 INTRODUCTION

E-learning is increasingly important in higher education, at centers for continuing education, and in the private sector. E-learning has mainly been realized with virtual learning environments (VLEs) in which a large number of online tests is carried out each year. For these online tests, various educational tasks are used with different didactical purposes. Frequently applied task types include multiple or single choice questions, matching tasks, or true/false questions, etc. Because many learners look at these tasks, their visual design is a key factor. User-centered design studies commonly focus on aesthetic aspects of the VLE. Underlying fundamental design principles in terms of didactics or psychology are often taken for granted. Nevertheless, such principles are crucial for understanding students' visual behavior as well as their learning processes. Understanding eye movements during completion of different task types has the potential for providing unique insights into human cognitive processes. Surprisingly, there is very little eye tracking research in the context of e-learning, although eye tracking methodology is becoming increasingly popular within user-centered design. This paper contributes to the bridging of this gap by introducing high-level analyses of eye movements recorded during the completion of different tasks embedded in a VLE.

2 RELATED WORK

There is surprisingly little e-learning research on differences in the visual behavior of learners that might emerge from different educational tasks, although early eye tracking work has shown that tasks have a significant impact on visual attention. Prior e-learning studies have investigated visual attention when completing specific tasks, e.g., arithmetic. The majority of studies appear to focus on specific aspects of multimedia learning, such as cueing, presentation speed, effect of expertise, or specific relations between various media types. Several overviews and examples are available [1] [2] [3]. In this paper, beyond traditional eye movement metrics, transition matrices are used to compare the effect of task on the distribution of visual attention.

Eye movement analysis in the context of e-learning has mainly focused on fixation counts and fixation durations for the evaluation of the utility of diagrams or animations accompanying text. Early well-known work includes Hegarty and Just's [4] study of learning from diagrams and verbal descriptions of pulley systems where learners tended to inspect the diagram at the ends of sentences and clauses. This suggests learners tended to fully interpret a sentence or clause before inspecting the referent in the diagram, referred to by others as text-directed behavior [5]. Although Hegarty and Just referred to this viewing behavior as highly interleaved, this early work lacked analysis of gaze transition between Areas Of Interest (AOIs), precluding subsequent statistical exploration. Schmidt-Weigand et al. [5] suggest that text-directed behavior might not be appropriate when visualizations accompanying text are dynamic. In their study of gaze-tracked visual attention distribution in a multimedia learning environment, the proportion of viewing time was analyzed in terms of fixations detected by a dispersion-based algorithm. Gaze transitions were also examined, defined as shifts of fixation from written text to visualization or vice versa. Transition matrices offer similar analysis, but at a finer level of detail, where transitions are tabulated among multiple AOIs. Lowe and Boucheix [6] echo prior concerns about animations in e-learning systems, suggesting that animations may be a double-edged sword for learners. They compare attention distribution in terms of fixation durations between static and dynamic segments of the stimulus, highlighting the sophisticated nature of cognitive processing of animations. They also note the need for a better understanding of the various perceptual and cognitive activities that learning involves. Providing statistical analysis of gaze switching behavior among AOIs, transition matrices can bolster this effort. More recent work employing AOI-based eye gaze analysis includes cued retrospective reporting [7], evaluation of instructional settings [8], and depictions of expert visual search strategies [9]. Although the proportion of fixations within AOIs is sometimes evaluated, transition matrices would augment this type of analysis with statistics of attentional switching behavior.

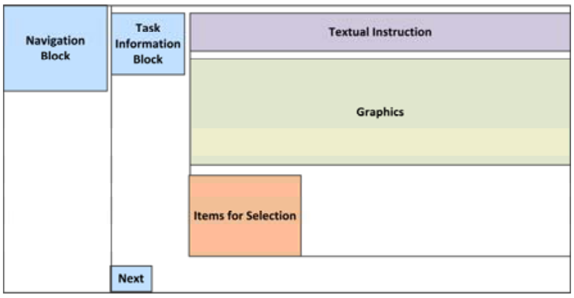
In similar prior work, Tsai at al. [10] tested multiple choice tasks in problem solving with three AOIs: selection items for answers, relevant as well as irrelevant details of the visual content, and the title (textual instruction). Results showed that, in general, learners focused mainly on those selection items that they regarded as the correct answer and fixated most on relevant details of the visual content. Fixating on eventually selected items may have been an instance of the gaze cascade effect [18]. However, transition matrices of the experiment did not yield any significant differences except for low-scoring students, who had repeatedly re-read textual instructions and shifted their visual attention to irrelevant details of the visual content. Their work focused mainly on the evaluation of differences between successful and unsuccessful students and did not necessarily consider the effect of task.

3 DESCRPTION OF THE STUDY

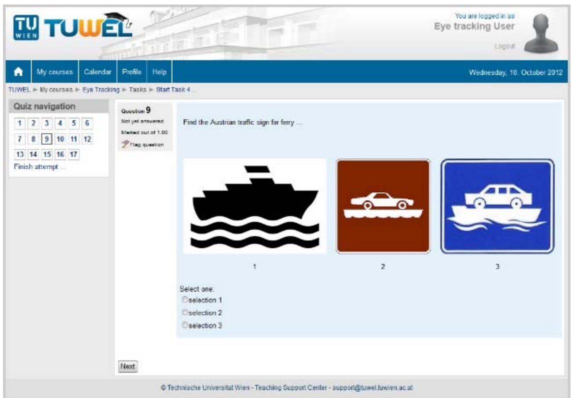
The present experiment is based on an earlier study where visual attention was investigated when learners completed a test involving foreign and domestic traffic signs. The experiment's test questions were embedded in an online e-learning course and learners from three different countries (Austria, China and the United States) answered various test questions about traffic signage. Earlier work investigated cultural and linguistic aspects and evaluated effects of training [11]. As the user interface of the VLE was identical for all participants across conditions, in this paper we ignore cultural and linguistic aspects and focus rather on the evaluation of the interface design. Analysis of students' eye movements is conducted from viewing of four different layout areas of the VLE, namely the area for textual instruction, visual content, items for selection, and items for navigation. Three tasks performed on the same VLE layout are evaluated: matching, true/false, and multiple choice. The analysis considers two types of gaze metrics. First, traditional eye tracking metrics are examined separately for each of the four layouts, grouped by task type. The research question addressed by this analysis is whether learners perceived different areas differently on each task. Second, gaze transition matrices for each task were generated in order to investigate the sequential patterns of eye movements. Transition matrices identify areas that are more likely transitioned to given a particular task. The research question addressed by this analysis is whether the task had an effect on distribution of visual attention.

3.2 Methodology

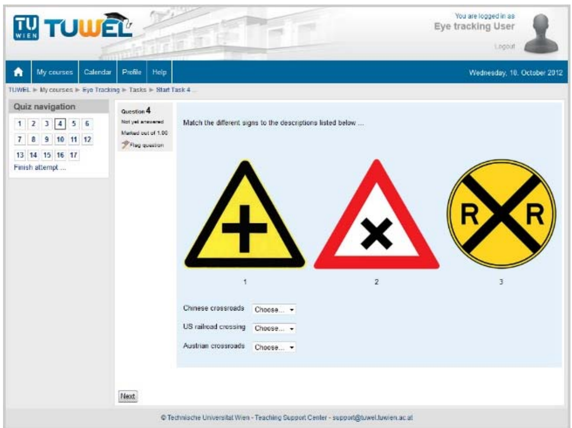
In order to answer the first research question the four layout areas of the VLE were defined in terms of AOIs (see Figure 1a). The analysis was designed as a single-user test scenario with one within-subjects factor, namely task type (at 3 levels: matching task, true/false task, or multiple choice task). In order to investigate differences of various eye tracking metrics within the four AOIs, one-way analysis of variance (ANOVA) was conducted for each condition. Post-hoc tests included pairwise t-tests with pooled SD and Bonferroni correction (Holm method for adjustment). To answer the second research question, transitions between the four layout areas were generated. For this the sequential order of fixations for all participants was retrieved and compared separately for each task type. Transitions include both saccades from one AOI to another, as well as consecutive saccades within an AOI.



(a) General Layout



(b) Multiple Choice Task



(c) Matching Task

(d) True/False Task

Figure 1. General Layout and Examples of each Task Type

3.3 Stimulus

All task types were embedded in an online course of the VLE. The VLE was based on Moodle, a widely-used open-source learning management system. Moodle's "Quiz" module was applied to set up the online test including a total of 35 test questions which all participants had to answer. Test questions consisted of 18 multiple choice tasks, 9 matching tasks and 8 true/false tasks. In multiple choice tasks (depicted in Figure 1b), learners had to find corresponding traffic signs indicated by the textual instruction. In matching tasks (see in Figure 1c) users had to match graphics to short descriptions as instructed in the textual instructions. In true/false tasks (shown in Figure 1d) users had to make a true/false decision about the textual instruction describing the visual content. The order of task types was counterbalanced via Latin square and test questions were displayed in random order. Learning tasks were selected based on high usage compared to other available question types. The goal of the e-learning course was to familiarize participants with foreign signage emulating driving education systems.

The four AOIs covered different semantic parts of the test questions (see Figure 1a). The first AOI covered *textual instructions,* which contained writing. In matching tasks, instructions prompted students to perform a certain action without containing any reference to the test content, e.g., *"Match the different images to the descriptions listed below".* The textual instructions of true/false tasks provided a clear statement with regard to content, but omitted any directions for action, e.g., *"The traffic sign prohibits turning left"*. Multiple choice task instructions included directions for action and asked students questions in regard to content, e.g., *"Find the traffic sign for yield"*. The second AOI covered the *visual content* of the test questions depicting the traffic signs (labeled "Graphics" in Figure 1a). The third AOI covered different *items of selection*. Multiple choice tasks used radio buttons corresponding to the number of graphics available (see Figure 1b). True/false tasks used two radio buttons, one each for the correct and incorrect answers (see Figure 1d). Matching tasks used dropdown elements consisting of numbers which had to be matched to the short descriptions (see Figure 1c). The fourth AOI covered *elements for navigation*, such as the button to proceed to the next question, or two blocks providing the status of the current question and the entire test (see Figure 1a).

3.4 Apparatus

For recording learners' eye movements, the Tobii X50 stand-alone and the Tobii ET-1750 eye trackers were used. The resolution of the screens was set for both devices to 1280×1024 pixels. Both eye trackers are binocular and sample at a rate of 50 Hz. No chin rest was used for either device.

3.5 Participants

Eye movements of a total of 36 participants were recorded. Students completing the online test were aged between 22 and 34 years (M=26.3). Gender balance was ensured as 18 males and 18 females were recruited for the eye tracking study. All participants were screened for any major visual deficiencies as well as for any other remarkable constraints concerning their physical condition. To be admitted to the study, volunteers had to have a driving permit for at least 2 years.

3.6 Eye Tracking Metrics

Gaze metrics were calculated based on fixations that were detected with filters defined by Tobii. The velocity-based filter was used, with velocity and duration thresholds set to 50 degrees/sec and 100ms. Five widely-used eye tracking metrics were evaluated, namely *total gaze duration*, *time to first fixation*, *mean fixation duration*, *fixation count* as well as *regression rate*. *Total gaze duration* is the total time in seconds in an AOI. In general, longer gaze durations indicate higher complexity of the visual stimulus and suggest higher cognitive effort for task completion. Longer gaze durations suggest greater user interaction due to the complexity of the GUI. *Time to first fixation* is a latency measure giving the time in seconds from the onset of a stimulus until the viewer's gaze in the AOI. This metric suggests search efficiency and localization potential of visual attention, often suggesting saliency of visual elements. The third metric *mean fixation duration* indicates the average duration of fixations within an AOI. Basically, long fixation durations suggest greater cognitive load when extracting a GUI element's meaning. A higher level of expertise, lack of interest, or lower stimulus complexity tend to elicit shorter mean fixation durations. *Fixation count* provides the total number of fixations within an AOI, basically indicating areas of semantic importance. Finally, the *regression rate* indicates the number of revisits to an AOI. This parameter generally indicates ambiguity of visual elements, or the participant's willingness to extract information [11] [12] [13].

3.7 Transition Matrices

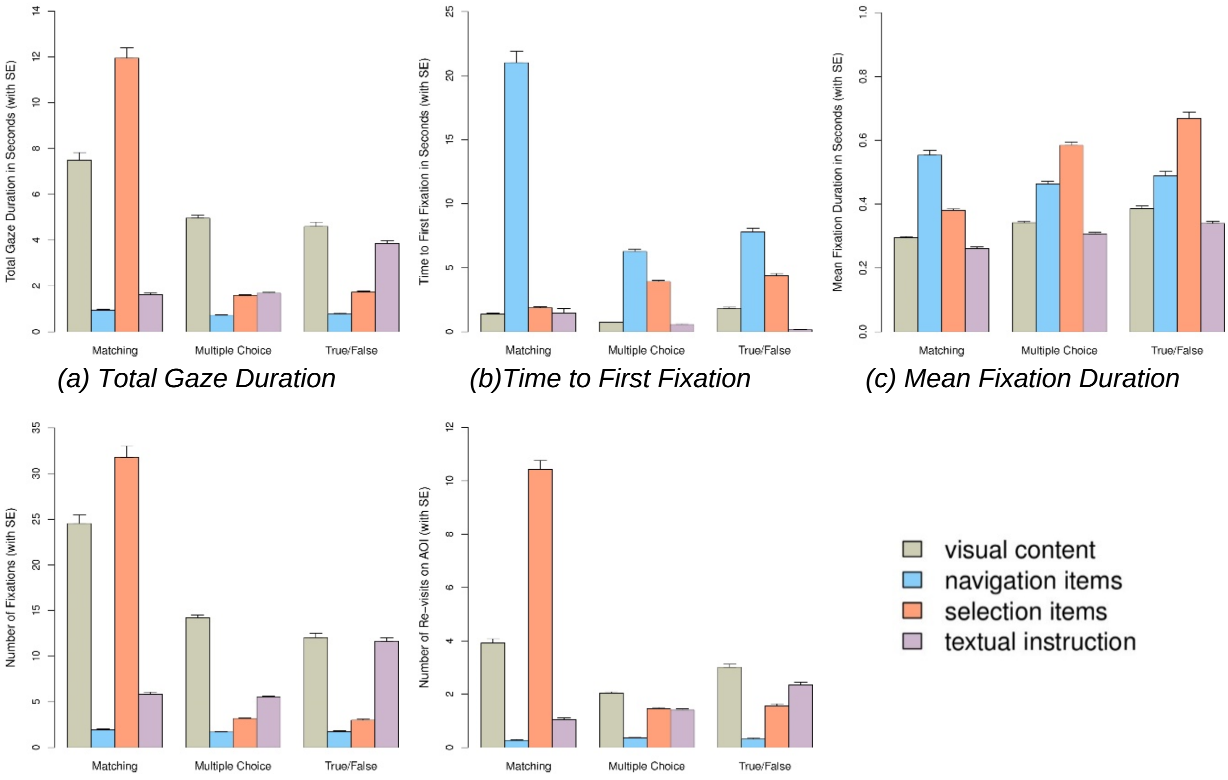
A transition matrix, introduced by Ponsoda et al. [14], is a tool for sequential gaze pattern analysis but for which statistical similarity measures are scarce. Although Ponsoda et al. used Z and X^2 statistics to compare matrices, their matrices were limited to cardinal (compass) saccade directions (i.e., N, NE, SE, etc.). In this paper, we use a method of computing and comparing transition matrices for any number of AOIs, based on Krejtz et al.'s [15] Markov model. The use of entropy resembles Goldberg and Kotval's [16] suggested use of matrix density.

A transition matrix is computed in the statistical analysis tool R for each of the AOIs defined in the stimulus image. Matrix elements are set to the number of transitions from each source AOI to each destination AOI for all participants. The matrix is then normalized relative to each source AOI (i.e., per row), such that each cell represents the estimated probability of transitioning from any AOI to any other given the first as the starting point. To compare the effect of task on gaze transitions, a statistical comparison of transition matrices is desired. To do so, empirical entropy H (with Miller-Madow correction), is computed via the bias-corrected maximum likelihood method as implemented by R's publicly available *entropy* module [17]. To facilitate statistical comparison of mean entropies per condition, H is computed per participant per condition. Therefore, a transition matrix is computed as above but per individual participant and per condition. This results in a table of entropies for each of the experimental conditions and each of the participants. ANOVA is then used to test for differences in mean entropy per condition.

4 RESULTS

4.1 Eye Movement Metrics Results

Figure 2 depicts the findings for all layout areas and for each task type, displayed separately for each eye tracking metric. *Total gaze duration* in AOIs yielded a significant effect of task. Focusing on textual instructions, the true/false task elicited the highest gaze durations (*$F(2,1212)=253.95$, $p<0.01$*). Gaze duration on visual content and items of selection differed significantly between tasks (*$F(2,1258)=51.54$, $p<0.01$; $F(2,1240)=825.48$, $p<0.01$, resp.*). Gaze durations on navigation items also differed significantly between tasks (*$F(2,1260)=14.38$, $p<0.01$*).



(a) Total Gaze Duration *(b)Time to First Fixation* *(c) Mean Fixation Duration*

(d) Fixation Count *(e) Regression Rate*

Figure 2. Results for all Eye Tracking Metrics (grouped by Task Type)

Similarly, *time to first fixation* revealed a significant effect of task. For textual instruction, the true/false tasks showed fastest entry times (*$F(2,1212)=12.89$, $p<0.01$*). In contrast, visual content was fixated last on the same true/false task (*$F(2,1258)=102.5$, $p<0.01$*). For selection items, GUI elements were fixated first on the matching task (*$F(2,1240)=82.97$, $p<0.01$*). Meanwhile, navigation items were fixated much later on the matching task than on other task types (*$F(2,1259)=332.99$, $p<0.01$*).

Mean fixation durations differed significantly across tasks. The matching task elicited the shortest durations while the true/false task yielded the highest mean fixation durations on all AOIs except for the navigation items. Differences for all AOIs were significant: textual instructions (*$F(2,1212)=31.99$,*

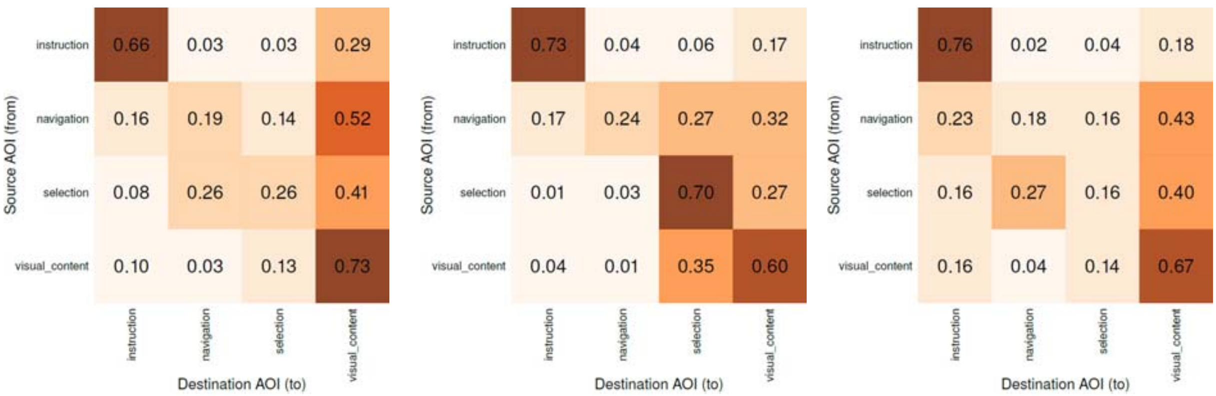
$p<0.01$), visual content ($F(2,1257)=63.69$, $p<0.01$), selection items ($F(2,1240)=98.31$, $p<0.01$), and navigation items ($F(2,1198)=15.95$, $p<0.01$).

When looking at *fixation count,* task yielded significant differences within AOIs. Textual instructions of true/false tasks garnered longer gaze durations than on other tasks ($F(2,1212)=209.73$, $p<0.01$). The fixation count on visual content was highest on the matching task ($F(2,1258)=130.7$, $p<0.01$). Selection items drew significantly larger fixation counts on the matching task ($F(2,1240)=900.92$, $p<0.01$). Navigation items revealed only marginally significant differences ($F(2,1259)=3.08$, $p<0.05$).

When focusing on *regression rate,* textual instructions were revisited most often in the true/false task ($F(2,1212)=68.18$, $p<0.01$). In contrast, visual content drew the highest number of regressions in the matching task ($F(2,1257)=105.12$, $p<0.01$). The matching task also elicited the largest number of regressions on selection items ($F(2,1240)=1012.7$, $p<0.01$). Task yielded no significant differences in fixation count on navigation items ($F(2,1198)=2.37$, *n.s.*).

4.2 Transition Matrix Results

A single transition matrix was constructed for each of the tasks. Each matrix, therefore, gives a visual snapshot of the probability distribution of gaze transition over the given AOIs. Matrix cells with higher probabilities indicate a higher (empirical) probability of transitioning to that (column) AOI given a fixation at the (row) source AOI for the given the task (see Figure 3).



(a) Multiple Choice *(b) Matching Task* *(c) True/False Task*

Figure 3. Gaze Transition per Task Type

An ANOVA (type 3) of mean transition matrix entropies indicates that, on average, gaze transitions differed significantly ($F(2,70)=28.18$, $p<0.01$). Pairwise t-tests (with Bonferroni correction) indicate that the mean entropy of the transition matrix for the matching task ($H = 1.85$, $SD= 0.12$) was significantly lower ($p<0.01$) than the mean entropies of the transition matrices for the multiple choice ($H=1.99$, $SD=0.11$) and true/false ($H=2.04$, $SD=0.18$) tasks. The difference between the latter two entropies was not significant.

5 DISCUSSION: INTERPRETATION OF GAZE ANALYTICS

When focusing on selection items, analysis revealed that dropdown elements were more visually demanding in matching tasks. In particular, the total gaze duration, fixation count, and regression rate were all much higher in the matching task than for true/false and multiple choice tasks, indicating differences in complexity and semantic importance (see Figure 4). This effect is bolstered by the transition matrix analysis, since the matching task elicited significantly more transitions to the selection item AOI (compare the selection AOI columns in Figure 3). Results also indicate that when performing the matching task, selection items tend to hold learners’ visual attention. Evidence for this is seen in the transition matrix with selection items being more likely to be targeted when looking at the selection items. This is also supported by the high regression rate metric. Furthermore, the time to first fixation revealed that dropdown items elicited early fixations. To sum up, dropdown items appear to be more visually salient, demanding to work with, and tend to attract visual attention early on.

Results for textual instructions showed specific patterns in eye movements for those instructions with greater reference to content. Simply put, the more information the instructions contained, the more they were likely to attract the learner's visual attention. Heatmaps (see Figure 4) show this effect qualitatively. The instruction of the true/false task drew the highest number of fixation counts in contrast to the other tasks.

Multiple Choice Task **Matching Task** **True/False Task**

(a) Heatmaps for each Task Types

(b) Gazeplot for each Task Type

Figure 4. Heatmaps and Representative Gazeplots for each Task Type

Comparison of mean fixation durations yielded a significant effect of task. Except for navigation items, results suggest that, for all AOIs, the true/false task elicited the highest average fixation durations (Figure 3c). This indicates higher cognitive load associated with the true/false task's decision making demands.

Transition matrices provide interesting insights into leaners' visual behavior. First, they show that transitions mostly occurred to textual instructions and visual content, and usually from the same AOI (i.e., a regression). This is evident in all tasks except the matching tasks where regression to the selection items AOI stands out. Secondly, when considering the matrix diagonal as an indicator of maintaining fixations atop the current AOI (e.g., regression), the matching task's matrix shows the learners' greater willingness to remain within the boundaries of the semantic layout on this task. For true/false and matching tasks transitions to other AOIs are more frequent, on the whole, as indicated by the entropy measures. That is, learners seem to process AOIs more conservatively (fewer transitions) on the matching task than on other tasks. Thirdly, transition matrices reveal a low probability for transitions from selection items back to the textual instructions. It may be that once learners perceived the selection items they tended to avoid revisiting textual instructions. This seems to occur for all task types and can be seen qualitatively in the exemplary scanpaths (see Figure 4b).

6 CONCLUSION & DESIGN RECOMMENDATIONS

The layout of an online e-learning test was compared in terms of eye movements when learners completed three different educational tasks. Visual behavior was analyzed along five traditional eye tracking metrics and a relatively new evaluation of gaze transitions. Comparison of transition matrices facilitated effects analysis between the three different tasks. Results clearly show that there are significant differences in visual behavior when performing different tasks. The following recommendations can be made from the findings of this study:

1. Dropdown elements appear to divert learners' visual attention away from other content such as instruction, and may be more cognitively demanding than radio buttons.
2. Textual instructions attract visual attention early on. Designers should consider placement of instructions at prominent areas in order to support learners' visual demands.
3. Compared to matching and multiple choice tasks, true/false tasks may be more cognitively demanding due to the underlying decision-making processes involved, which may impact visual processing. VLE designers should strive to balance cognitive demands of differing tasks.

To sum up, results of this study imply that e-learning designers have to be aware that the selection of different task types influence learners' visual perception, which may impact learning outcomes.

REFERENCES

[1] Van Gog, T.; Scheiter, K. (2010) Eye tracking as a tool to study and enhance multimedia learning. In: *Learning and Instruction*, Vol. 20, No.2, pp.95-99.

[2] Dogusoy, B.; Cagiltay, K. (2009) An Innovative Way of Understanding Learning Processes: Eye Tracking. In: *Human-Computer Interaction. Interacting in Various Application Domains, Lecture Notes in Computer Science*, Vol. 5613, pp.94-100, Springer Verlag, Berlin Heidelberg.

[3] Hyönä, J. (2010) The use of eye movements in the study of multimedia learning. In: *Learning and Instruction*, Vol. 20, No.2, pp.172-176.

[4] Hegarty, M.; Just, M.A. (1993) Constructing Mental Models of Machines from Text and Diagrams. In: *Journal of Memory and Language*, Vol.32, pp.717-742.

[5] Schmidt-Weigand, F.; Kohnert, A.; Glowalla, U. (2010) A closer look at split visual attention in system and self-paced instruction in multimedia learning. In: *Learning and Instruction*, Vol.20, No.2, pp.100-110.

[6] Lowe, R.; Boucheix, J.M. (2011) Cueing complex animations: Does direction of attention foster learning processes? In: *Learning and Instruction*, Vol.21, No.5, pp.650-663.

[7] De Koning, B.B.; Tabbers, H.K.; Rikers, R.M.J.P.; Paas, F. (2010) Attention guidance in learning from a complex animation: Seeing is understanding? In: *Learning and Instruction*, Vol.20, No.2, pp.111-122.

[8] Jarodzka, H.; Scheiter, K.; Gerjets, P.; van Gog, T. (2010) In the eyes of the beholder: How experts and novices interpret dynamic stimuli. In: *Learning and Instruction*, Vol.20, pp.146-154.

[9] Vitak, S.A.; Ingram, J.E.; Duchowski, A.T.; Ellis, S.; Gramopadhye, A.K. (2012) Gaze-Augmented Think-Aloud as an Aid to Learning. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1253-1262, ACM, New York.

[10] Tsai, M.J.; Hou, H.T.; Lai, M.L.; Liu, W.Y.; Yang, F.Y. (2012) Visual attention for solving multiple-choice science problem: An eye-tracking analysis. In: *Computers & Education*, Vol.58, No.1, pp.375-385.

[11] Rakoczi, G.; Duchowski, A.; Casas-Tost, H.; Pohl, M. (2013) Visual perception of international traffic signs: influence of e-learning and culture on eye movements. In: *Proceedings of the 2013 Conference on Eye Tracking South Africa*, pp. 8-16, ACM, New York.

[12] Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H. (2011) *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford University Press, Oxford.

[13] Goldberg, J.H.; Wichansky, A.M. (2003) Eye tracking in usability evaluation: A practitioner's guide. In: *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, Hyönä, J., Radach, R., Deubel, H., Eds., pp.493-516, Elsevier Science, Amsterdam.

[14] Ponsoda, V.; Scott, D.; Findlay, J.M. (1995) A probability vector and transition matrix analysis of eye movements during visual search. In: *Acta Psychologica*, Vol.88, pp.167-185.

[15] Krejtz, K.; Szmidt, T.; Duchowski, A.T.; Krejtz, I. (2014) Entropy-based Statistical Analysis of Eye Movement Transitions. In: *Proceedings of the 2014 Symposium on Eye-Tracking Research and Applications,* ACM, New York.

[16] Goldberg, J.H.; Kotval, X.P. (1999) Computer Interface Evaluation Using Eye Movements: Methods and Constructs. In: *International Journal of Industrial Ergonomics,* Vol.24, pp.631-645.

[17] Hausser, J.; Strimmer, K. (2009) Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks. In: *Journal of Machine Learning Research*, Vol.10, pp.1469–1484.

[18] Shimojo, S; Simion, C.; Shimojo, E.; Scheier, C. (2003) Gaze bias both reflects and influences preference. In: Nature Neuroscience, Vol.6, No.12,pp.1317-1322.