Eye Movement Patterns in Solving Science Ordering Problems

Hui Tang

University of Georgia

Lisa Kendhammer

University of Georgia

Scott A. Brown

University of Georgia

Elizabeth Day

University of Georgia

James N. Moore

University of Georgia

Norbert J. Pienta

University of Georgia

Dynamic biological processes, such as intracellular signaling pathways, commonly are taught in science courses using static representations of individual steps in the pathway. As a result, students often memorize these steps for examination purposes, but fail to appreciate either the cascade nature of the pathway. In this study, we compared eye movement patterns for students who correctly ordered the components of an important pathway responsible for vasoconstriction against those who did not. Similarly, we compared the patterns of students who learned the material using three dimensional (3-D) animations previously associated with improved student understanding of this pathway against those who learned the material using static images extracted from those animations. For two of the three ordering problems, students with higher scores had shorter total fixation duration when ordering the components and spent less time (fixating) in the planning and solving phases of the problem-solving process. This finding was supported by the scanpath patterns that demonstrated that students who correctly solved the problems used more efficient problem-solving strategies.

**Keywords: Eye-tracking, problem-solving, scanpath, science ordering problems**

Introduction

It is common practice in the biological sciences for students to learn dynamic processes, such as intracellular biochemical cascades, using a series of static images that represent the steps in a process. The apparently static nature of these processes is reinforced when students refer to textbooks wherein the information is presented using simplistic diagrams in which components of a pathway are connected by arrows. Unfortunately, many students spend a great deal of time memorizing the steps in the process for an examination, but fail to fully appreciate that the cascade itself is critical for the process to occur or the spatial relationships among the components of the cascade (i.e., where in a cell the steps occur and how they are interrelated).

To address these problems, investigators have created and assessed the effectiveness of three dimensional (3-D) animations that depict the dynamic nature of these processes (Buchanan et al, 2005; Reindl et al, 2015). In the former study, not only were the students’ scores for content accuracy significantly higher in the course taught using the animations, but their scores for understanding the pathway as a cascade and the spatial organization of the components involved in the pathway were increased to the same extent.

Received December 15, 2015; Published February 28, 2016.

Citation: Groner, R., Bischof, W. F. & Siegenthaler, E. (2016). How to Prepare your Manuscript for the Journal of Eye Movement Research. Journal of Eye Movement Research, x(y):z, 1-3.

Digital Object Identifier: 10.16910/jemr.x.y.z

ISSN: 1995-8692

This article is licensed under a [Creative Commons Attribution 4.0 International license. ](https://creativecommons.org/licenses/by/4.0/)

In order for students in the biological sciences to envision how specific alterations in intracellular pathways might cause adverse effects in certain diseases, they must be able to mentally reconstruct the pathway. There are two approaches to assess students’ understanding of these processes on examinations. One is to present the components of the pathway in various orders, and task the students with recognizing which is correct. The limitation to this approach is that a student simply must identify one component in the incorrect position to eliminate that choice. The more difficult method, which more accurately assesses the students’ mastery of the material, is to present the components of the pathway in random order and have the students rearrange them in the correct order.

An alternative way to learn more about how students solve ordering problems such as these, is to combine a computer program that allows students to “drag and drop” the components of the pathway into their individually selected “correct order” with eye- tracking technology to capture this information. In this way, the participant’s series of steps can be analyzed to reveal information about how each participant went about solving the problem. Using this approach, investigators could be informed about the students’ thought process as they rearrange the components in the pathway.

In recent years, eye-tracking has been used to investigate the cognitive activities of students as they solve problems in science courses. Results of such studies may provide valuable information about students’ expertise, problem-solving strategies and cognitive processes, and may contribute to curriculum improvement and assessment development (Tai, Loehr & Brigham, 2006). In this study, we explored the differences in eye movement patterns between students who correctly ordered the components of the alpha-adrenergic pathway and those who did not. To determine whether or not the manner in which the material was presented to them impacted their ability to correctly answer these questions, we compared the responses of students who learned the material using 3-D animations previously used by Buchanan et al (2005) against those of students who viewed static images taken from those animations.

*Cognitive Theory of Multimedia Learning*

Mayer’s cognitive theory of multimedia learning guides the design of many educational materials (Mayer & Moreno, 2002). In this theory, working memory is composed of a visual sketchpad processing center and phonological/verbal loop processing center. Information processed by both centers (via multiple modalities of input) is more effective than by one processing center alone. Therefore, use of both centers allows extraneous cognitive load to be reduced, and has led to several instructional design principles (Mayer, 2005). For example, Mayer suggested that using animation and narration together could help students learn effectively (Mayer & Sims, 1994).

*Animations in Biological/Medical Education*

The use of animations in biological and medical science education is a popular topic of study (O'Day, 2007; Tversky, Morrison & Betrancourt, 2002; Ruiz, Cook & Levinson, 2009; Tunuguntla, Rodriguez, Ruiz, Qadri, Mintzer, Van Zuilen & Roos, 2008; Gookin, McWhorter, Vaden & Posner, 2010). A literature review on the use of animations in medical education found that studies comparing the impact of animations against static graphics for different topics yielded conflicting results (Ruiz et al., 2009). For example, in teaching and learning science concepts, animations either surpassed (Tversky et al., 2002) or were as effective as static images (Tunuguntla et al., 2008). In other studies, animations could be distracting or even interfere with the construction of knowledge and were more difficult to understand by novices (Gookin et al., 2010). These conflicting results were due to various factors, including cognitive load introduced by presentations of the information and participants’ prior knowledge.

*Eye-tracking and Problem-solving*

Eye-tracking research is based on the eye-mind hypothesis, which states that eye movement is the observable measure of visual attention that is linked to the cognitive processing of information (Just & Carpenter, 1984; Eivazi & Bednarik, 2011; Doherty, O’Brien & Carl, 2010). The use of eye-tracking to study problem-solving in science and mathematics education has emerged as a viable technique in the last three decades (Suppes, Cohen, Laddaga, Anliker & Floyd, 1983; De Corte, Verschaffel & Pauwels, 1990; Hegarty, Mayer & Green, 1992; Epelboim & Suppes, 2001; Grant & Spivey, 2003; Green, Lemaire & Dufau, 2007; Tang & Pienta, 2012; Tang, Kirk & Pienta, 2014). The majority of these studies have utilized multiple-choice problems that can be solved by choosing the correct answer with a simple mouse click. The design and data analysis of such studies are uncomplicated, and the experimental procedure is straightforward and short. However, the simplistic nature of multiple-choice items requires the use of complementary experimental methods like verbal protocols to support or qualify the researchers’ conclusions.

To obtain insight into participants’ cognitive processes, researchers have utilized more complex formats of items including arithmetic equations, graphs, animations or simulations. In many of these studies, participants verbalized their solutions during eye-tracking (concurrent “think aloud”) or post-experiment interviews (retrospective “think aloud”) (van Gog, Paas & van Merriënboer, 2005). Nevertheless, transcribing and coding verbal data are time-consuming and costly (Chi, 1997; Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka & Van de Weijer, 2011, p. 292), and “think aloud” interviews may influence participants’ behaviors and/or mental workload in certain settings (Hertzum, Hansen & Andersen, 2009).

Therefore, a problem type that can reveal participants’ cognitive processes with relatively simple eye-tracking data should overcome the above shortcomings. Ordering problems (also called sequencing or drag-and-drop problems) task students with placing a randomized set of items or steps into their correct order. This type of problem is ideal for investigations of the students’ understanding of the interdependency of individual steps in a biological process, as was the focus of the present study. It also permits assessment of the students’ cognitive load during problem-solving because the number of items or steps involved may be directly related to number of bits of information in working memory. It is also ideal for an eye-tracking experiment, because participants can click and move the mouse to rearrange the items on a computer screen, thereby eliminating the need for participants to “think aloud” or write on computer screens while their step-wise approach to the solution (i.e., problem-solving strategy) is recorded.

In the present study, we conducted an eye-tracking experiment to investigate undergraduate students’ eye movement patterns and corresponding cognitive activities as they solved ordering problems in physiology. To understand how students solve problems, some researchers divide problem-solving processes into sub-phases. Researchers have found that eye movement patterns in the same phases can be different between two groups of participants, or can vary in different phases within a group (Tang & Pienta, 2012; Tang et al., 2014; Hegarty et al., 1992). Depending on the characteristics of the task being performed, the phases and their components have been proposed differently (Hegarty et al., 1992; De Corte et al., 1990; Green et al., 2007; Mayer, Larkin & Kadane, 1984). In this study, we divided the problem-solving process into three phases: *reading-and*-*planning*, *problem-solving* and *answer-checking*. *Reading-and*-*planning* is defined from the time when a participant began to read a problem to the time before the participant completely moved a choice into a step box. *Problem-solving* is defined from the end of *reading-and*-*planning* to the time when the participant moved the last choice into a step box and did not make any subsequent changes. The third phase, *answer-checking,* is from the end of *problem-solving* to the time when the participant clicked the “submit” button.

*Eye Movement Measures*

Eye-trackers provide a variety of eye movement measures. Holmqvist et al. (2011) proposed a taxonomy for eye movement measures that included movement, position, numerosity, latency, and distance. The authors considered over 100 measures in these four categories. Eye-tracking studies typically use two basic types of measures: fixation and saccade. Scanpath, a third type of eye movement measure, also is used (Poole & Ball, 2005). A fixation is defined as “the state when the eye remains over a period of time”. A saccade is “the rapid motion of the eye from one fixation to another” (Holmqvist et al., 2011, p. 21-23). A scanpath is simply constructed from the fixations and saccades using temporal sequencing.

In eye-tracking research, fixation duration may be the most-frequently reported eye movement measure; other common measures include fixation count and visit count (Holmqvist et al., 2011, p. 377, 413, 417). Fixation count and visit count also are referred to as number of fixations and number of dwells, respectively (Holmqvist et al., 2011, p. 412-417). According to the Tobii Studio 2.X manual (Tobii, 2010), fixation duration (previously called “fixation length”) measures “the duration of each individual fixation within an area of interest (AOI)”; fixation count “measures the number of times the participant fixates on an AOI or an AOI group”; and visit count (previously called “observation count”) “measures the number of visits within an AOI or AOI group”. Because one visual visit on an AOI may consist of multiple fixations, visit count is always smaller than or equal to fixation count within an AOI.

It is generally accepted that fixation duration and fixation count are correlated to difficulty, complexity, importance and interest of visual materials or tasks, which reflect the cognitive load of the tasks (Slykhuis, Wiebe & Annetta, 2005). Specifically, longer fixation duration corresponds to higher complexity of visual materials; larger fixation number indicates more importance of the region being perceived (Hegarty, Mayerm & Green, 1992; Green, Lemaire & Dufau, 2007). Furthermore, scanpath reflects viewing strategies, which imply processes of cognition (Slykhuis et al., 2005; Peebles & Cheng, 2001). As a result, researchers use scanpath patterns to examine characteristics of one’s cognitive process and to differentiate problem-solving strategies between groups of participants. Studies have shown that experts and novices have different scanpath patterns when they view materials and solve problems (Tai, et al., 2006; van Gog, et al., 2009a; Tang, Topczewski, Topczewski & Pienta, 2012). For example, experts more quickly find and look relatively longer at relevant information, and thus solve problems more effectively than novices (Tang et al., 2012; Tsai et al., 2012).

Purpose of the Study

The purpose of the present study was to explore the relationship between students’ success in solving science ordering problems and factors recorded during eye-tracking experiments. The intent was to correlate eye movement measures during the problem-solving process. The authors also were interested in possible correlations between the eye movement measures themselves, and the scanpaths of students who solved the problems correctly versus those of students who did so incorrectly. Finally, the effects of media type (dynamic animations vs. static images) on student performance were examined. The research questions were:

1. What are the correlations between fixation duration, fixation count, visit count, time on task and number of mouse clicks?

2. What are the relationships between student performance and the above factors?

3. What are the differences in scanpath patterns between students who solved the problems correctly and incorrectly?

4. How does media type affect whether students solve the problems correctly or incorrectly?

Methodology

*Participants*

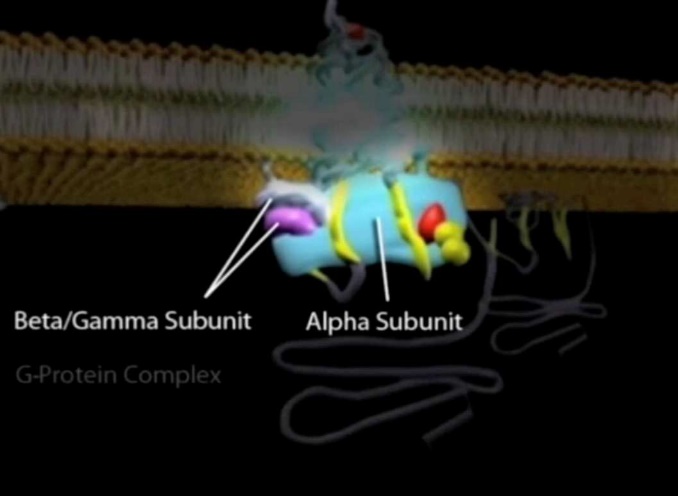
The participants were students enrolled in an undergraduate physiology course at a large Southeastern comprehensive university during the spring 2014 semester. A total of 89 students volunteered to participate in the Institutional Review Board (IRB) approved study and received compensation for their time. Approximately 80% of the participants had previously taken a biochemistry course that covered the alpha-receptor pathway used as the subject of this study. The participants were randomly assigned to two groups that differed by the type of educational media they viewed -- dynamic animations versus static images. Out of the 89 participants, data from eight were discarded due to poor quality of eye-tracking recordings, which was based on the cutoff value (80%) of the eye-tracking ratio (Kruger, Hefer & Matthew, 2014). Consequently, 81 participants consisting of 31 male participants and 50 female participants remained in the analysis. Among the 81 participants, 39 watched the media containing dynamic animations and 42 watched the media containing static images.

*Apparatus*

Participants’ eye movements were monitored and recorded using a Tobii T120 eye tracker with a sampling rate of 60 Hz. The threshold for defining a fixation was set to be 100 ms with a radius of 30 pixels, with an accuracy of 0.5 degrees. The display of the eye tracker included a 17-inch LCD screen with a resolution of 1280 × 1024 pixels. The data were recorded and processed using Tobii Studio 2.0.4 software (Tobii, 2010).

*Materials*

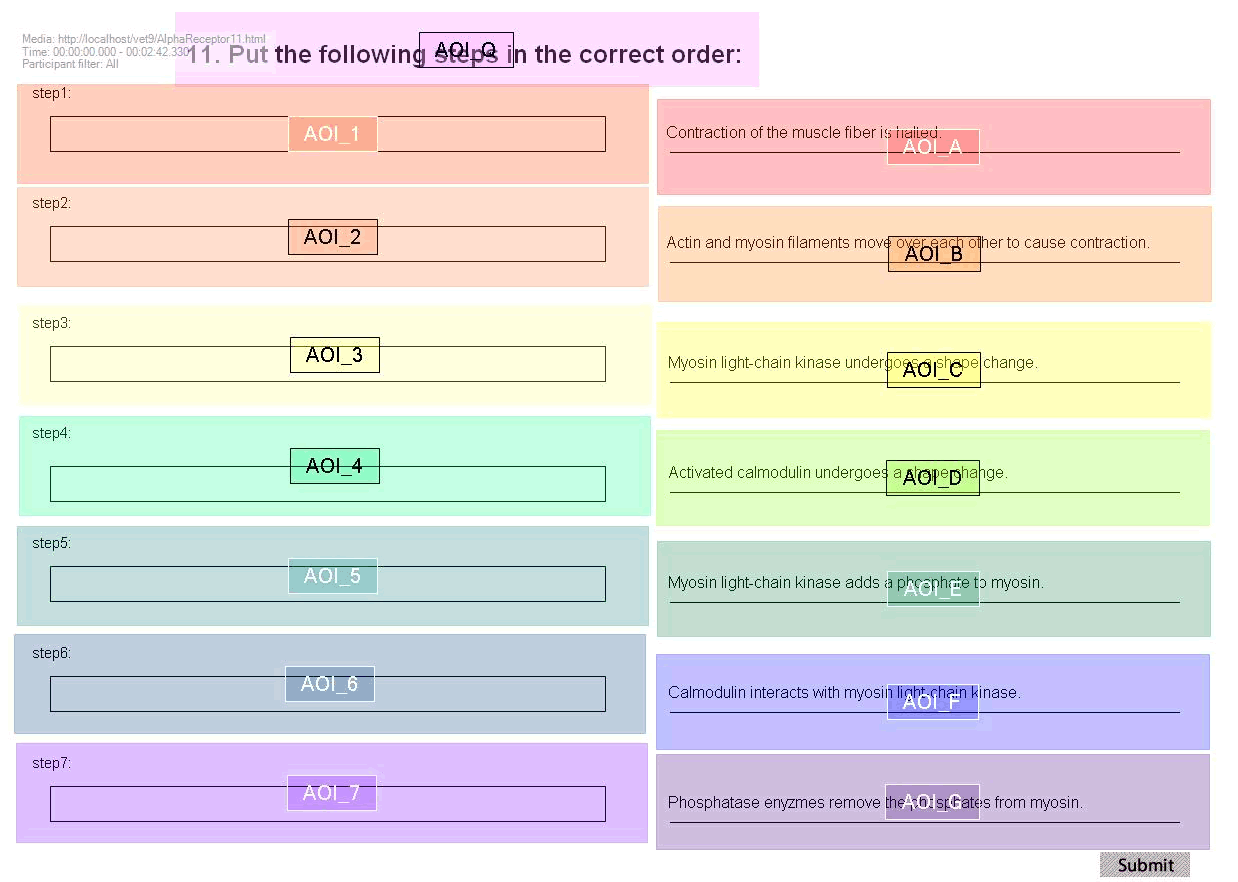
The educational media used in the experiment, which simulated the process of the alpha receptor pathway, was designed by the authors. The two types of media, dynamic animation and static images, had the same length (~5 min) and content, including the same narration. The animations were those used in the study by Buchanan et al. (2005) and depicted the dynamic interactions of the processes. The static images displayed static representations, which were 30 screenshots from the dynamic animations. Figure 1 shows a representative static image.



*Figure 1. Screenshot of the media showing components of the alpha receptor pathway.*

After watching the media, the participants were presented 17 problems to solve, which were in the same order. These problems were developed by the authors and were available on the intranet hosted by the same server where the eye tracker was connected. Each problem was displayed on a corresponding website with a submit button. Two types of problems were designed, one was multiple-choice (14 problems) and the other was ordering (3 problems). Each of the 3 ordering problems covered a specific component of the alpha receptor pathway and was considered to be of equal difficulty by the two physiology professors who created the problems. Two postdoctoral physiology researchers and two graduate students reviewed and solved the problems independently before the experiment. As their answers were in 100% agreement, they were used as the answer key to grade the participants’ responses.

In this paper, we report the analyses and the results for the ordering problems. Each of the ordering problems included 6 to 7 sub-steps of the physiologic pathway that the participants had learned about by watching either the media containing the dynamic animations or static images. The participants were randomly assigned a media type before he/she began the study. In the instructions for each ordering problem, the participants were tasked with putting the steps in the correct order (Figure 2). The websites on which the problems were presented allowed the participants to move their choices from the right side of the screen to ordered step boxes on the left side (i.e., drag-and-drop). If a choice had already been moved to a specific step box, it could be moved to a new step box. In either case, when a participant moved a choice into a step box, the initial location of that choice, either above the line on the right side or in the box on the left side, became empty. If the participant dragged a choice out of a step box and released the mouse without moving the choice into a new step box, the choice automatically returned to its original location.



*Figure 2. Superimposition of area of interest (AOI) fields on the screen capture of the third ordering problem. AOIs 1 through 7 are for the steps (left side), AOIs A through G are for the choices (right side), and AOI\_Q is for question stem.*

*Procedure*

After reading the instructions and signing the consent form, the participants sat in front of the eye tracker screen at a distance of 60−65 cm. They were required to sit as still as possible during the experiment to minimize gaze drift. Once the eye tracker was calibrated, participants watched their assigned media twice. After the instructor stopped the eye tracker, the participants were asked to relax for 1 minute before the eye tracker was recalibrated. Finally, the participants were instructed to solve the problems at their own paces. The experiment lasted 30 to 50 minutes, depending on the participants.

Data Analysis

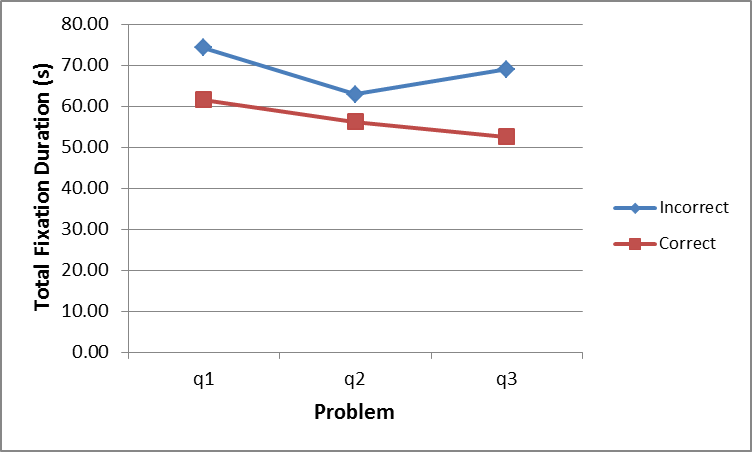
AOIs were defined for each choice, step box, and the problem stem: letters for choices/problem stem and digits for steps (Figure 2). The first approach used for data analysis in this study involved using logistic regressions to explore relationships between how well the participants solved the problems and a set of independent variables. The dependent variable was the participants’ total scores earned for solving the three problems. A student earned 1 point if the order of a problem was completely correct; otherwise, 0 points were assigned. Thus, a participant’s total score could be 0, 1, 2, or 3 for all the three problems. We also used another grading system in which the grade was based on the Levenshtein distance (Tang et al., 2012) between a participant’s answer and the answer key. For example, the answer key of Problem 3 is DFCEBGA (Figure 5). If a participant placed the order as DFCEBAG, the Levenshtein distance was 2 and the grade was 7 – 2 = 5. Thus, in this grading system, partial credits were assigned if part of the step order in a problem was placed correctly. The independent variables included media type, total fixation duration (sum of all the fixation durations within an AOI or AOI groups), fixation count, visit count, time on task and number of mouse clicks (hereafter *mouse clicks*). In the analysis, fixations on each problem (i.e., sum of AOIs) rather than on each AOI were examined. However, there were still 18 independent variables for the three problems. Based on previous studies, it was likely that some of these variables could be correlated, particularly the eye movement measures (Stieff & Hegarty, 2011; Tang & Pienta, 2012; Tang et al., 2014; Williamson et al., 2013; Jang et al., 2011). Therefore, we explored the correlations between these variables and expected the dimension of data to be reduced based on the results. All the analyses were performed using R 3.1.2 (R Core Team, 2015).

The second approach used for data analysis was a comparison of the eye movement patterns between the participants who solved the problems correctly (“correct group”) and those who solved them incorrectly (“incorrect group”). This analysis was conducted using a software tool developed by our research team, which had the similar functions of eyePatterns (West et al., 2006) and was written in R. In the software tool, a scanpath pattern is a subsequence with at least three AOIs that presents in a certain number of the scanpaths. In this study, we used “collapsed” scanpaths and 20% was used as the cut-off to define a scanpath pattern. That is, a subsequence with at least three AOIs that was shared by at least 20% of the scanpaths in a group was considered to be a pattern. For example, if there are 30 scanpaths in a group and a subsequence with three AOIs can be detected in six of the scanpaths, then this subsequence is a pattern of this group. If no pattern was identified in a group using the 20% cutoff, a lower cutoff (e.g., 15% or 10%) was employed. A group of scanpaths usually has multiple patterns and each scanpath in the group may contain none, one or more than one pattern.

Results

*Problems*

The numbers (%) of participants who solved problems 1, 2 and 3 correctly were 46 (56.8%), 9 (11.1%) and 47 (58.0%), respectively. All of the correlations (Pearson’s *r*) between the variables (total fixation duration, fixation count, visit count, time on task and mouse clicks) within the same problem were larger than 0.56, indicating large effect sizes according to Cohen’s criteria (Cohen, 1988). Specifically, total fixation duration, fixation count and time on task were highly correlated (*r* ≥ 0.95). The correlation matrix is presented in Appendix. As a result, total fixation duration was used for the subsequent analyses because it is one of the most commonly-reported eye movement measures and it could represent the other variables. The average total fixation durations for students who solved each problem correctly and incorrectly are shown in Figure 3.



*Figure 3. Average* *total fixation durations in each problem.*

Results of a logistic regression performed to explore the relationship between the total score and total fixation duration in each problem as well as media type are presented in Table 1. An ordered logistic regression was performed rather than a binomial logistic regression because the dependent variable (score) had more than two possible values (Venables & Ripley, 2002): 0, 1, 2, and 3 (see Data Analysis).

The total fixation durations in Problems 1 and 3 between the correct and incorrect groups were significantly different (*p* < 0.05). The negative coefficients indicated that participants with lower scores fixated longer on the problem. The total fixation durations in Problem 2 between the two groups were not significantly different, nor were media types. Because of the strong correlations identified previously between the five measures, the findings for total fixation duration should apply equally to fixation count, visit count, time on task, and mouse clicks.

*Table 1*

*Logistic regression results showing relationship between students’ scores and fixation duration in each problem as well as media type*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Coefficient** | **Std. Error** | **t value** | **p value** |
| Media Type | 0.161 | 0.424 | 0.379 | 0.705 |
| Problem 1 | -0.017 | 0.008 | -2.026 | 0.043\* |
| Problem 2 | 0.013 | 0.011 | 1.233 | 0.217 |
| Problem 3 | -0.023 | 0.010 | -2.266 | 0.023\* |

\* *p* < 0.05

*Phases*

The average total fixation duration in each problem-solving phase and the percentage of each phase in a problem are listed in Table 2. The total fixation durations in Phase 1 for each problem had a narrow range (17.40s ± 0.45s), as did the percentages of Phase 2 durations (62.32% ± 1.07%). Because nine independent variables (3 phases × 3 problems) and media type remained, dimension reduction was again considered before the logistic regression was conducted.

*Table 2. Average total* *fixation duration (in seconds) in each phase and the percentage in a problem*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Problem** | **Group** | **Phase 1** | **Phase 2** | **Phase 3** | **Sum** |
|  | incorrect | 18.51 (24.92%) | 50.38 (67.82%) | 5.39 (7.26%) | 74.29 |
| 1 | correct | 15.77 (25.62%) | 37.87 (61.54%) | 7.90 (12.84%) | 61.54 |
|  | average | 16.95 (25.83%) | 43.28 (63.39%) | 6.82 (10.78%) | 67.04 (100%) |
|  | incorrect | 17.64 (28.06%) | 39.82 (63.32%) | 5.42 (8.62%) | 62.89 |
| 2 | correct | 19.45 (34.62%) | 33.04 (58.80%) | 3.70 (6.58%) | 56.18 |
|  | average | 17.84 (31.14%) | 39.07 (61.25%) | 5.23 (7.61%) | 62.14 (100%) |
|  | incorrect | 17.25 (25.00%) | 49.34 (71.53%) | 2.39 (3.47%) | 68.98 |
| 3 | correct | 17.59 (33.51%) | 31.63 (60.27%) | 3.26 (6.22%) | 52.49 |
|  | average | 17.44 (31.80%) | 39.07 (63.12%) | 2.90 (5.08%) | 59.41 (100%) |

This confirms the findings in Table 1 that in general, total fixation duration in each phase was not dependent on number of steps in problem. Therefore, the average total fixation duration of the three problems in each phase was included in the subsequent logistic regression.

The results of the ordered logistic regression, presented in Table 3, indicate that as a whole, participants with lower scores fixated significantly longer in Phase 2, whereas participants with higher scores fixated significantly longer in Phase 3.

Using the Levenshtein distance system in which partial credits were assigned to the part of the step order that was placed correctly, the results were similar to those obtained (Tables 1 and 3) when 1 point was assigned only if the step order in a problem was completely correct.

*Table 3*

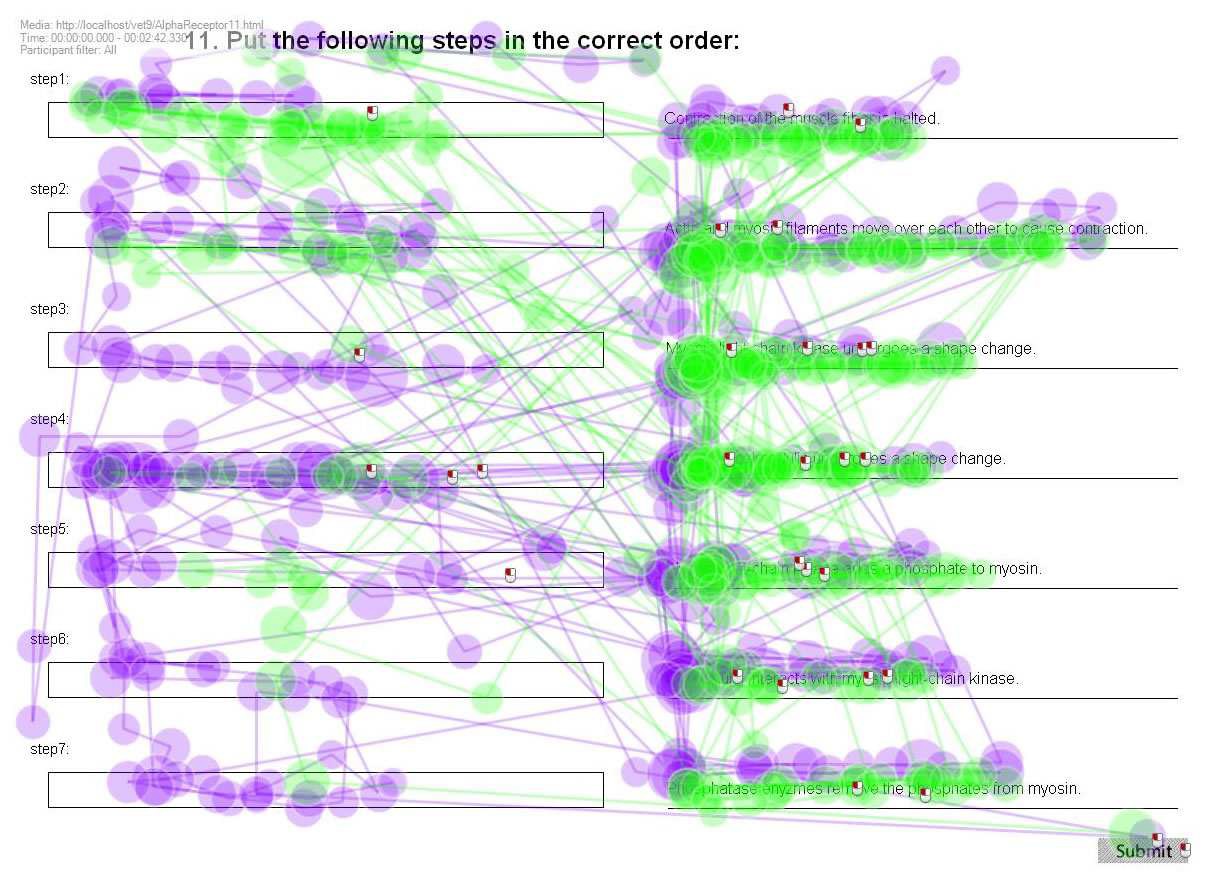
*Logistic regression results showing relationship between student score and total* *fixation duration in each phase*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Coefficient** | **Std. Error** | **t value** | **p value** |
| Media Type | 0.024 | 0.433 | 0.055 | 0.956 |
| Phase1 | 0.000 | 0.030 | -0.012 | 0.991 |
| Phase2 | -0.061 | 0.015 | -3.999 | 0.000\*\* |
| Phase3 | 0.097 | 0.047 | 2.050 | 0.040\* |

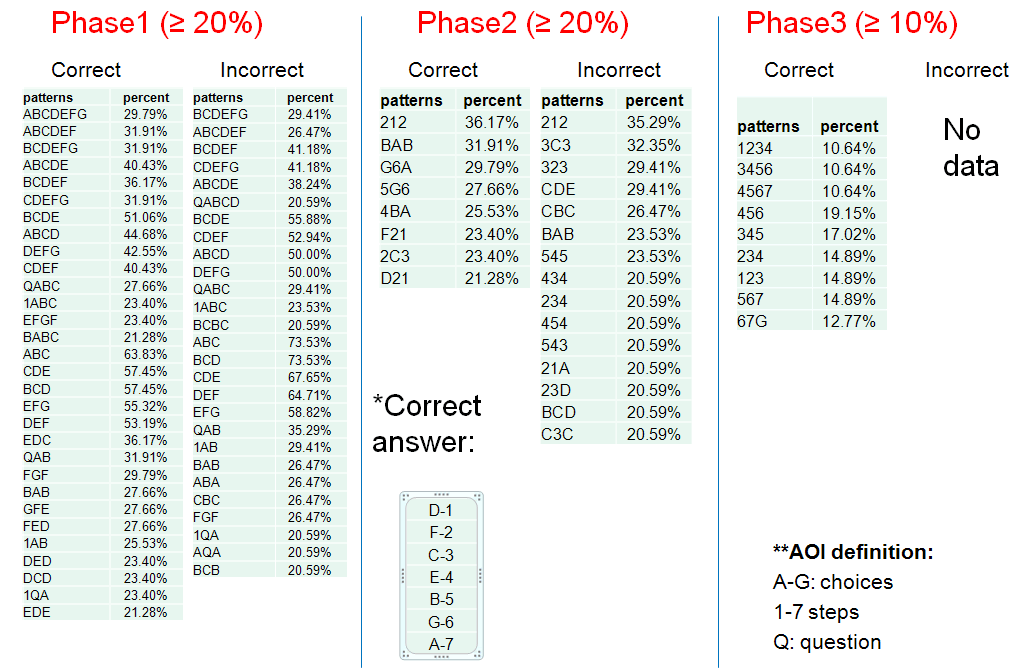
\* *p* < 0.05

*Scanpath patterns*

As an example, the scanpaths (gaze plots) for Problem 3 from two participants (one correct and one incorrect, in green and purple, respectively) are shown in Figure 4. The difference of gaze plots on the *step* side can be observed: for the correct participant, the fixation counts generally decrease from step 1 to step 7; for the incorrect participant, gazes were roughly distributed on each step. To illustrate the details of scanpaths, the patterns in each phase of students who solved Problem 3 correctly and incorrectly were listed side by side (Figure 5).



*Figure 4. Gaze plots of the third ordering problem from a student who solved the problem correctly (green) and student who solved incorrectly (purple).*



*Figure 5. Scanpath patterns in each phase of the third ordering problem for students who solved the problem correctly and incorrectly. For each group (correct or incorrect), column 1 contains the scanpath pattern and column 2 contains the percentage of the occurrence of pattern, i.e., the ratio of the number of participant who had a pattern to the total number of participants in a group. The definition of AOIs is listed at the bottom of Phase 3, and the correct order of the choices (answer key) in Problem 3 is listed at the bottom of Phase 2. For example, choice D should be placed in step box 1, choice F in step box 2, choice C in step box 3, etc.*

As is evident in Phase 1 (Figure 5), the correct and incorrect groups had similar scanpath patterns. Their eye movements were mostly between adjacent choices, e.g., DCD or FED, without any jumps (e.g., no patterns such as ADE were identified). There were also a few patterns involving the problem stem (e.g., QABCD) and the first step (e.g., 1ABC).

In Phase 2, the correct and incorrect groups had different scanpath patterns. While the eye movements in the incorrect group were still between adjacent choices and/or steps such as 545 and 21A, most patterns in the correct group revealed eye movements between choices and the corresponding correct steps such as F21 and 5G6 (both F-2 and G-6 are correct choice-step connections). We can use the ratio of fixational transitions to further illustrate the pattern difference between the two groups. In this study, the ratio of fixational transitions is defined as the number of fixational transitions within the areas of choices (e.g., 21) and steps (e.g., BA) relative to the number of transitions between them (e.g., 5G). When computed from the patterns in Figure 5, the ratio was 7:9 for the correct participant group and 24:6 for the incorrect group. The latter had much more within fixational transitions and less between transitions.

In Phase 3, neither group had patterns shared by at least 20% of the participants. When 10% was used as the cutoff value, the scanpath patterns in the correct group revealed that participants checked their answers step-by-step. There was no scanpath pattern identified in the incorrect group, suggesting that participants who solved the problem incorrectly either did not check their answers or their eye fixations were largely scattered if they performed any “answer-checking.” The latter is supported by the previous finding that the participants who ordered the steps in a problem correctly fixated longer on Phase 3 of the problem, than the participants who ordered the steps incorrectly.

Discussion

The results of the present study revealed that students with higher scores fixated for a significantly shorter amount of time on the choice in the *problem-solving* phase (Phase 2), but fixated longer in the *answer-checking* phase (Phase 3). We also used another grading system in which a student was assigned partial credits if part of the step order in a problem was placed correctly. The results were similar to those where a student was assigned 1 point only if the step order in a problem was completely correct. The differences in total fixation duration between the correct and incorrect groups can be explained by the results obtained after comparing the scanpaths. In Phase 2, most eye movement patterns in the correct group were connections between choices and the corresponding correct step boxes, which was evidence of an efficient problem-solving strategy. In contrast, the eye fixations of participants in the incorrect group wandered across adjacent choices and/or step boxes (e.g., 545 and 21A). In Phase 3, participants in the correct group checked their answers in order, whereas there was no evidence that participants in the incorrect group checked their answers. The results showed that *answer-checking* was more likely an indicator of better outcome/performance in problem-solving or a test.

In Phase 1, there was no significant difference in total fixation duration between the two groups, and the scanpath analysis revealed that both groups had very similar eye movement patterns (i.e., primarily between neighboring choices, e.g., DCD or FED). Thus, Phase 1 was actually a “*reading*” stage, which was also supported by the scanpath patterns that revealed normal reading orders (e.g., QABCD and 1ABC). In other words, before completely moving a choice into a step box, none of the participants formulated a “plan” that could be detected by eye-tracking. Therefore, it appears that “*planning*” was integrated with “*problem-solving*” in Phase 2. As summarized in Table 2, the total fixation durations in Phase 2 on different choices were not different for participants in the two groups. This finding is consistent with those indentified in our previous studies regarding the reading phase (Tang & Pienta, 2012; Tang et al., 2014). Thus, if students are at a similar academic level (e.g., enrolled in the same course), their total fixation durations in the reading phase are relatively constant when they solve a science problem. The durations may only vary according to the length of problem, regardless of whether they ultimately solve the problem correctly or not.

There is evidence in the literature that dynamic animations are more effective in aiding student learning than static images (Buchanan et al, 2005; Tversky et al., 2002; Aldahmash & Abraham, 2009). However, the results of the logistic regressions in this study indicated that media type did not positively impact students’ scores. There are two explanations for why no significant difference was discovered, with the first being that some students might have had prior knowledge in this topic area. Approximately 80% of the participants had previously taken a biochemistry course in which the pathway used in this study could have been covered. If most of the participants had prior knowledge in this content area, we would not expect to see a large difference due to media type. The fact that >88% of the students answered Problem 2 incorrectly would suggest that prior exposure to the pathway did not, however, equate with a full understanding of the materials. The second reason for the lack of a difference due to media type could be that the static images themselves were taken from the dynamic animations and were quite detailed. Given how comprehensive these static images were, it is possible there was not enough difference between the static images and the dynamic animations to discover a significant difference between the participants who watched each media type.

In this study, we determined that total fixation duration, fixation count, visit count, time on task and mouse clicks were positively correlated with each other. This result provides additional reliable evidence supporting similar findings regarding fixation duration and fixation count detected in our previous research studies (Tang & Pienta, 2012; Tang et al., 2014) and by others in studies on problem-solving (Stieff & Hegarty, 2011; Williamson et al., 2013; Jang et al., 2011). Because studies addressing this issue included different problem formats (e.g., multiple-choice, word problems and ordering problems) and diverse disciplines in science education (e.g., chemistry and physiology), it appears that at least in science problem-solving research, (total) fixation duration and fixation count are interchangeable. In many eye-tracking studies, researchers have reported multiple eye movement measures simultaneously. For example, in the fields of psychology and reading, where more than one experiment usually is included in a study, fixation duration and fixation count are often analyzed and reported in each of the experiments, though the results on these two measures could be very similar. The results from the present study suggest that correlations between different eye movement measures should be examined to avoid repetitive analyses.

Moreover, researchers have used eye-tracking data to predict viewing (Kanan, Ray, Bseiso, Hsiao & Cottrell, 2014; Greene, Liu & Wolfe, 2012) or problem-solving behaviors and outcomes (Bednarik, Eivazi & Vrzakova, 2013; French & Thibaut, 2014; Eivazi & Bednarik, 2011; Tsai, Viirre, Strychacz, Chase & Jung, 2007). In most predicting models, multiple eye movement measures are selected as predictor variables. These measures include fixation duration, number of fixations, fixation angles and saccade amplitude. According to the findings in the literature and the current study, some of these measures may be highly correlated. Therefore, researchers are advised to consider collinearity when building predictive models. If some eye movement measures are highly correlated, it is appropriate to remove variables from the models (Dormann et al., 2013).

Limitations and Future Work

In Phase 2, the scanpaths were generated directly from the eye-tracking data. Thus, an AOI represented eye fixations on the original location of a choice or step box. For example, when AOI = A, the participants could be looking at the location of the first choice, regardless of whether the choice was still present or had been moved (although the possibility that latter scenario happened was very low). Our concern would be similar for AOI = 1, for which our scanpath data did not inform us as to which choice was in the first step box where the participants were fixating; or whether the participants were fixating on an empty step box. Thus, although the scanpath patterns in this phase revealed logic connections between choices and the corresponding correct steps (in the correct group), combining both eye-tracking and action recordings should provide more accurate information. Moreover, in the present study, eye movement patterns were compared mainly between two groups of participants. In the future, we also plan to compare more within-subject behaviors. For example, if a participant solves one problem correctly and solves another one incorrectly, are there differences in eye movement patterns between the two problem-solving processes?

Conclusions

The key finding of the current study was that the total fixation duration of students with higher scores was significantly shorter on 2 of the 3 problems than that of students with lower scores. By analyzing data for each phase of the problem-solving process, significant differences were detected during Phase 2, i.e., *planning* and *problem-solving*, which is consistent with the findings in our previous studies (Tang & Pienta, 2012), as well as during Phase 3, the *answer-checking* phase. In addition, the scanpath analysis revealed different eye movement patterns in these two phases of problem-solving between correct and incorrect groups of participants. These results indicate that ordering problems are appropriate types of activities for eye-tracking researchers to explore problem-solving strategies. Finally, no significant differences in eye movement measures were identified between students who watched the media featuring dynamic animations and static images.

References

Aldahmash, A. H., & Abraham, M. R. (2009). Kinetic versus static visuals for facilitating college students’ understanding of organic reaction mechanisms in chemistry. *Journal of Chemical Education,* *86*, 1442–1446.

Bednarik, R., Eivazi, S., & Vrzakova, H. (2013). A Computational Approach for Prediction of Problem-Solving Behavior Using Support Vector Machines and Eye-Tracking Data. In *Eye Gaze in Intelligent User Interfaces* (pp. 111-134). Springer London.

Buchanan, M. F., Carter, W. C., Cowgill, L. M., Hurley, D. J., Lewis, S. J., MacLeod, J. N., & Robertson, T. P. (2005). Using 3D animations to teach intracellular signal transduction mechanisms: taking the arrows out of cells. *Journal of veterinary medical education*, *32*(1), 72-78.

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The journal of the learning sciences*, *6*(3), 271-315.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology*, *82*(2), 359.

Djamasbi, S., Mehta, D., & Samani, A. (2012). Eye Movements, Perceptions, and Performance. In *Proceedings of the eighteenth Americas Conference on Information Systems (AMCIS)* (p. 1).

Doherty, S., O’Brien, S., & Carl, M. (2010). Eye tracking as an MT evaluation technique. *Machine translation*, *24*(1), 1-13.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., & Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46.

Eivazi, S., & Bednarik, R. (2011). Predicting problem-solving behavior and performance levels from visual attention data. In *Proc. Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI* (pp. 9-16).

Epelboim, J., & Suppes, P. (2001). A model of eye movements and visual working memory during problem solving in geometry. *Vision Research*, *41*(12), 1561-1574.

French, R. M., & Thibaut, J. P. Using eye-tracking to predict children's success or failure on analogy tasks. In *Livre/Conférence Proceedings of the Thirty-Sixth Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society* (pp. 2222-2227).

Gookin, J. L., McWhorter, D., Vaden, S., & Posner, L. (2010). Outcome assessment of a computer-animated model for learning about the regulation of glomerular filtration rate. *Advances in physiology education*, *34*(2), 97-105.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving guiding attention guides thought. *Psychological Science*, *14*(5), 462-466.

Green, H. J., Lemaire, P., & Dufau, S. (2007). Eye movement correlates of younger and older adults’ strategies for complex addition. *Acta psychologica*, *125*(3), 257-278.

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers’ task from eye movement patterns. *Vision research*, *62*, 1-8.

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

Hertzum, M., Hansen, K. D., & Andersen, H. H. (2009). Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, *28*(2), 165-181.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press. Chapter 9.

Jang, Y. M., Lee, S., Mallipeddi, R., Kwak, H. W., & Lee, M. (2011, January). Recognition of human’s implicit intention based on an eyeball movement pattern analysis. In *Neural Information Processing* (pp. 138-145). Springer Berlin Heidelberg.

Just, M. A, & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), New Methods in Reading Comprehension Research (pp. 151-182). Hillsdale, NJ: Erlbaum.

Kanan, C., Ray, N. A., Bseiso, D. N., Hsiao, J. H., & Cottrell, G. W. (2014, March). Predicting an observer's task using multi-fixation pattern analysis. In *Proceedings of the symposium on eye tracking research and applications* (pp. 287-290). ACM.

Kruger, J., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. *L2 Journal of Eye Movement Research*, *7*(5), 4.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational psychology review*, *14*(1), 87-99.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology*, *86*(3), 389.

Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 231-273). Hillsdale, NJ: Erlbaum.

O'day, D. H. (2007). The value of animations in biology teaching: a study of long-term memory retention. *CBE-Life Sciences Education*, *6*(3), 217-223.

Peebles, D., & Cheng, P. C. H. (2001). Graph-based reasoning: From task analysis to cognitive explanation. In *Proceedings of the twenty-third annual conference of the cognitive science society* (pp. 762-767).

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reindl, K.M., White, A.R., Johnson, C., Bender, B., Slator, B.M., & McLean, P. (2015). The virtual cell animation collection: Tools for teaching molecular and cellular biology. PLoS Biol 13(4): e1002118, 1-9.

Ruiz, J. G., Cook, D. A., & Levinson, A. J. (2009). Computer animations in medical education: a critical literature review. *Medical education*, *43*(9), 838-846.

Slykhuis, D. A., Wiebe, E. N., & Annetta, L. A. (2005). Eye-tracking students' attention to PowerPoint photographs in a science education setting. *Journal of Science Education and Technology*, *14*(5-6), 509-520.

Stieff, M.; Hegarty, M.; Deslongchamps, G. Identifying Representational Competence with Multi-representational Displays. *Cognition Instruct.* 2011, *29* (1), 123−145.

Strandvall, T. (2009). Eye tracking in human-computer interaction and usability research. In *Human-Computer Interaction–INTERACT 2009* (pp. 936-937). Springer Berlin Heidelberg.

Suppes, P., Cohen, M., Laddaga, R., Anliker, J., & Floyd, R. (1983). A procedural theory of eye movements in doing arithmetic. *Journal of Mathematical Psychology*, *27*(4), 341-369.

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 assessments. *International journal of research & method in education*, *29*(2), 185-208.

Tang, H., & Pienta, N. (2012). Eye-tracking study of complexity in gas law problems. *Journal of Chemical Education*, *89*(8), 988-994.

Tang, H., Kirk, J., & Pienta, N. J. (2014). Investigating the Effect of Complexity Factors in Stoichiometry Problems Using Logistic Regression and Eye Tracking. *Journal of Chemical Education*, *91*(7), 969-975.

Tang, H., Topczewski, J. J., Topczewski, A. M., & Pienta, N. J. (2012, March). Permutation test for groups of scanpaths using normalized Levenshtein distances and application in NMR questions. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 169-172). ACM.

Tobii. (2011). *Tobii Studio 2.X User Manual.* Stockholm, Sweden.

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine*, *78*(Supplement 1), B176-B185.

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. *Computers & Education*, *58*(1), 375-385.

Tunuguntla, R., Rodriguez, O., Ruiz, J. G., Qadri, S. S., Mintzer, M. J., Van Zuilen, M. H., & Roos, B. A. (2008). Computer-based animations and static graphics as medical student aids in learning home safety assessment: a randomized controlled trial. *Medical teacher*, *30*(8), 815-817.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International journal of human-computer studies*, *57*(4), 247-262.

van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, *19*, 205-221.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer, New York.

West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006, March). eyePatterns: software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 symposium on Eye tracking research & applications* (pp. 149-154). ACM.

Williamson, V. M., Hegarty, M., Deslongchamps, G., Williamson III, K. C., & Shultz, M. J. (2013). Identifying student use of ball-and-stick images versus electrostatic potential map images via eye tracking. *Journal of Chemical Education*, *90*(2), 159-164.

Appendix

*Table A1*

*Correlation matrix of fixation duration, fixation count, visit count, time on task and mouse click in each of the three problems*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Problem  **Fixation** 1  2  **Duration** 3 | **Fixation Duration** | **Fixation Count** | **Visit Count** | **Time on Task** | **Mouse Clicks** |
| 1 2 3  0.308 0.399  0.430 | 1 2 3  **0.977** 0.308 0.411  0.263 **0.968** 0.441  0.344 0.389 **0.974** | 1 2 3  **0.856** 0.141 0.331  0.081 **0.766** 0.352  0.240 0.274 **0.878** | 1 2 3  **0.990** 0.302 0.406  0.272 **0.973** 0.382  0.377 0.425 **0.976** | 1 2 3  **0.663** -0.003 0.153  -0.089 **0.606** 0.185  0.095 0.138 **0.713** |
| **Fixation** 1  2  **Count** 3 |  | 0.303 0.389  0.445 | **0.896** 0.149 0.316  0.125 **0.809** 0.359  0.280 0.332 **0.903** | **0.976** 0.264 0.355  0.277 **0.949** 0.342  0.393 0.443 **0.950** | **0.677** -0.033 0.129  -0.083 **0.566** 0.154  0.118 0.156 **0.699** |
| **Visit** 1  2  **Count** 3 |  |  | 0.147 0.325 | **0.891** 0.127 0.289  0.154 **0.842** 0.307  0.330 0.402 **0.915** | **0.768** -0.064 0.171 |
|  | 0.425 | -0.102 **0.641** 0.211 |
|  |  | 0.127 0.206 **0.805** |
| **Time** 1  2  **Spent** 3 |  |  |  | 0.288 0.400  0.418 | **0.680** -0.016 0.149  -0.075 **0.605** 0.208  0.120 0.124 **0.735** |
| **Mouse** 1  2  **Click** 3 |  |  |  |  | -0.141 0.110  0.228 |

Red and bold: *r* ≥ 0.95. Bold: *r* > 0.56.

14