Visual Attention during E-Learning: Eye-tracking Shows that Making Salient Areas More Prominent Helps Learning in Online Tutors

Farnaz Tehranchi[#] (farnaz.tehranchi@psu.edu)
Frank E. Ritter[†] (frank.ritter@psu.edu)
Chungil Chae[†] (chadchae@gmail.com)
Department of Computer Science and Engineering
† College of Information Sciences and Technology
Penn State, University Park, PA 16802 USA

Abstract

In this study, we investigate how high- and low-performance learners (N=12) act differently while using a cognitive tutoring system. We examine three research questions: (1) Can we predict learners' performance using only their visual attention (eye movement data)? (2) Can we predict learners' performance from visual attention data and initial performance? (3) Are age, gender, first language, where they look, and the sequence of Areas of Interests (AOIs) significant factors in the learners' performance? Learners more correctly answer questions taken from larger rather than smaller AOIs. Our results show that high-performance learners pay more attention to the content that contains answers to later questions. Surprisingly, the tutor did not change the learners' visual search to a goal-oriented search. Our analyses can help instructional designers create a more productive learning experience because visual search behavior as part of a learner model with acceptable accuracy in early stages can be used in adaptive tutors. Additionally, we trained a classifier on the eye movement data to predict learners' performance for each question. Its results provide a list of suggestions for designing more productive learning experiences, such as enticing user attention by increasing the size of the content that contains answers and changing the order of contents.

Keywords: eye-tracking; eye movement data; learner modeling; e-learning; online tutoring system; cognition analysis; visual attention

Introduction

We present an experimental evaluation of visual attention information (i.e., gaze-based and eye movements data) as a source for modeling a user's learning in e-learning and to understand patterns of visual behavior. We started with these primary research questions: (a) How do learners interact with the materials in a tutor, including quiz questions? (b) What are the different ways learners interact with screen objects? (c) Where did learners who answered correctly and incorrectly look? (d) How often did learners spend time in interactive areas? (e) How much time did learners look at different objects? We used a cognitive tutor that was implementing the Declarative-to-Procedural (D2P) theory on a tutoring system. D2P is designed to support tutoring procedural skills that can be, and need to be, described to learners initially with declarative knowledge, and to support instructional designers. D2P draws on general theories of learning and provides multimedia instruction pages followed by multimodal quizzes to test and proceduralize the declarative knowledge (Ritter et al., 2013).

There are many technologies for investigating cognitive processes more deeply and accurately, one of which is eye tracking. Tracking people's eye movements can help us understand visual and display-based information processing and the factors that may impact the usability of system interfaces. Also, eye movement data advance learning interfaces and experiences by suggesting how to alter interfaces for improved learning (e.g., Worsley, Barel, Davison, Large, and Mwiti, 2018). In this way, eye-movement data can provide an objective source of interface-evaluation to inform the design of improved interfaces. Eye movements recorded by eye trackers can provide a lot of information on cognitive processes (e.g., Holmqvist et al., 2011; Salvucci and Goldberg, 2000).

Also, paying attention to how users look for information on online tutors is becoming increasingly important in designing positive learning experiences. Lai et al. (2013) review the effect of eye-tracking approaches in a study of learning based on more than 100 studies. These studies try to connect learning outcomes to cognitive processing. These studies mainly focused on the effect of individual differences rather than the effect of designs on the learning environment and patterns of visual behavior. Eye movement data give us a way to tie together the eyes and higher-level functions of the brain. Mapping eye-movement patterns to cognitive strategies help researchers understand the psychological causes of behavior (e.g., eye movement and retrieval process, Anderson, Bothell, and Douglass, 2004).

In our experiment, an eye tracker recorded learners' visual attention while they studied the tutor contents and when they answered questions. We also calculated learners' visual attention information such as eye fixation over Areas of Interest (AOIs) and normalized dwell (dwell time divided by AOI Coverage). AOIs help researchers analyze the various components of a visual scene.

We observe how long learners looked at the AOIs and the scan order (AOIs Sequence). These gaze data are suggested to be a good predictor of learning (Kardan & Conati, 2012). However, evaluating learners' performance based on learners' interaction data from e-learning environments is challenging (Holzinger, Kickmeier-Rust, Wassertheurer, & Hessinger, 2009). Additionally, the instructional designer's approach can be improved by eye gaze data (Clinton, Cooper,

Michaelis, Alibali, & Nathan, 2017). Furthermore, collecting interaction data, gaze data, in an open-ended environment is computationally expensive (Kardan & Conati, 2013). Therefore, for complex, dynamic, and interactive environments, researchers must define the relevant AOIs. The effect of most extended fixations in overall performance has been studied before with AOI-related and AOI-independent features (Bondareva et al., 2013). This demonstrated that attention defining the instructional materials is essential for assessing learning.

Objects' arrangements on a page alter information processing and patterns of visual behavior. Attention's order of AOIs has been studied; for instance, during program debugging—Lin et al.'s (2016) study noted that eye gaze data during debugging reveals a dynamic and nonlinear procedure for debugging. This procedure can be used to understand the cognition of information processing. In addition, the theory of visual hierarchy helps instructional designers avoid potential design problems (Faraday, 2000).

In light of the above, we attempt to predict performance in the tutor on a question-by-question basis using just gaze data analyzed based on AOIs. AOIs have been defined using the answers of questions in D2P content pages. This leads to the following questions:

Question 1: Can we predict performance by having only learner's visual attention (eye movement data)?

Question 2: Can we predict a later performance with current performance (first question set) and visual attention?

Question 3: Are age, gender, first language, where learner's look, and the sequence of AOIs significant factors in the performance? Do these factors interact with the eye-movement data when predicting performance?

Method

Participants

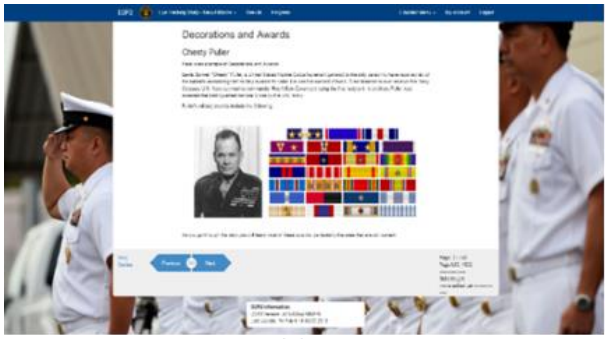
We collected data from 12 volunteer participants that were summer interns (graduate and undergraduate) or full-time employees at ACT Inc. ACT is an educational testing company that administers the ACT college preparatory test. Participants were unfamiliar with the tutoring system (D2P) and content (Navy ribbons).

Materials and apparatus

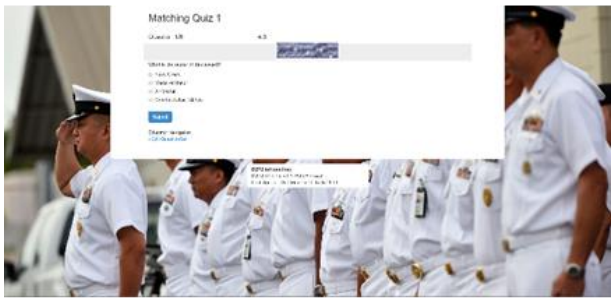
Declarative-to-procedural (D2P) tutoring system. We used a tutor created in the D2P tutoring system[1]. D2P is designed based on the learning and memory theories in Adaptive Control of Thought-Rational, ACT-R (Anderson, 2007). D2P is a page-based tutor. Content pages can include text, video, pictures, and simulations. An example of existing pages for contents and quizzes are shown in Figure 1.

We selected a tutor from the existing tutors created in D2P, one that teaches knowledge about US Navy Ribbons. The D2P/Navy Ribbon tutor provides an overview and the details of the ribbons (medals) awarded by the US Navy. This tutor is primarily designed to provide an example D2P tutor for testing the tutor architecture and, secondarily, to help recognize Navy ribbons. This tutor describes ribbon types, who is eligible to receive them, and for what they are awarded. Participants also learn how to recognize ribbons, whether they are currently awarded, special locations or situations medals were awarded for, and their precedence—that is, the order that they appear on a uniform. This tutor primarily uses pictures and text to describe the ribbons followed by quizzes and practice pages so that participants can practice recall and recognition of ribbons. The time on task and question answers are stored in the tutor's database.

Eye Tracker. We used a SMI Experiment SuiteTM 360° eye-tracker. It is a tracker bar with cameras attached to the bottom of the screen, and a 17" monitor driven by a PC. The tutor was displayed in an Internet Explorer browser.



(a)



(b)

Figure 1: (a) A content page and (b) a question with timer in the Navy Ribbon tutor in D2P.

Design and procedure

All participants navigated the tutor using a computer in an experiment room. After consent, learners were calibrated on the eye-tracker. Their gaze data were recorded with the eye tracker, and their performance was the number of correct answers on the 89 multiple choice quiz questions in the tutor. Participants used a mouse for all inputs.

The tutor contains five quizzes. Participants went through 13 content pages and answered 38 questions in the first quiz, the first set of questions. Then the participants went through 12 content pages and then answered the next 51 questions as

[1] http://acs.ist.psu.edu/projects/d2p/d2p.html

part of the second set of questions. Questions followed the same pattern in both sets and were associated with the same details of content pages. The question set 1 contains two quizzes, one about matching ribbons' name (AOI 1) and image (AOI 2), as shown in Figure 1b. The military decorations quiz features a series of questions about the first set of ribbons' information in the tutor, such as eligibility. Question set 2 contains three quizzes: a matching quiz, a military decorations quiz about the second set of ribbons, and a matching quiz about all ribbons that are in the first and second part of the tutor's content.

Analysis

We used the analysis software provided with the eye tracker, BeGaze[2], to aggregate the eye movements and saccades into the AOIs shown in Figure 2. The minimum fixation threshold was 50 ms, and fixations below this threshold were ignored.

We defined four AOIs based on the tutor questions and answers with different sizes: (a) AOI 1 is the name of the ribbon being tested, (b) AOI 2 is the image of the ribbon, (c) AOI 3 is the main ribbon information—Type, Eligibility, Awarded for, and Status, and (d) AOI 4/White is the rest of tutor's screen (anything else on the page not related to a question). Each AOI was adjusted to include its elements on each content page. Therefore, we knew which AOI provided the answer to each question. This varied by questions. Answers were only in AOIs 1 to 3, and not in AOI 4.

In the analysis, we combined two sets of features. The first set of features includes the learners' performance (or grade per question), age, first language, and gender. The second set of features consists of statistical measures of participants' attention and gaze information collected by the eye-tracker such as fixation time (time spent in each AOI), sequence (order of viewing AOIs), normalized dwell, and first fixation duration. Learners' behavior created a dataset with 1,864 data points with 42 features (15 distinct features for the classifier) that each data point contains the time and grade of that AOI per question and per person. We excluded the 4 pages where learners watched instructional videos and quiz pages that contain question sets. Features include eye tracker data collected for each page and AOI. In this study, the trial starts when participants enter the new page. Participants were free to navigate the tutor and used the previous and next buttons. Therefore, we had multiple trials for some pages (stimulus).

Exploratory data analysis

Because AOI 4/White corresponds to areas that did not hold answers to any questions, there were not any grades for these data points. Therefore, the AOI 4/White data points were removed from the analyses, as well as the 221 data points representing AOIs that are related to answers but not examined on a given page by a given subject. After the removal and data cleaning, there remain 1262 data points.

To start exploring the data, we were interested in seeing how well each participant did in each of the two sets of

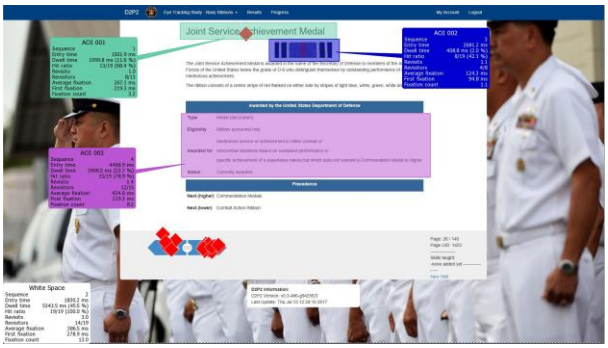


Figure 2: AOI 001, AOI 002, and AOI 003 (counterclockwise from upper left) are areas related to the question sets. White Space is the remaining screen area containing the navigation bar, buttons, and page details. Red dots are users' click events. BeGaze software provides statistical information for each AOIs. The BeGaze software generates the AOI labels (e.g., AOI 001).

questions. Figure 3a shows the percentage of correctly answered questions for each of the participants. Of the 12 participants, 9 had a higher percentage of correct answers in the second set than the first set. Figure 3b shows a boxplot of the percentage of correct answers for each of the two sets of questions. Overall, the median was lower in the first set than the second. It is also interesting to note that the spread of percent corrects decreased from the first set of questions to the second set.

Results

We first describe the descriptive statistics and summarize where the participants looked. Also, we examine the reliability of these measures and how they might be used to distinguish high- and low-performers. Next, we investigate how visual attention information might be able to predict learners' performance.

Descriptive statistics

All of the participants were able to perform the task and finish the tutor. The study took about 30~40 minutes. Participants spent additional time on non-content pages and tasks including question answering, video watching, and setting up. Figure 4a shows the time each participant spent on the AOI's. Figure 4b shows variability across individuals and pages. In Figure 4 participant P11, who has the lowest number of correct answers as shown in Figure 3, also had the lowest average fixation time in AOIs 1-3.

Figure 5 breaks down further the times per AOI per page. Some pages were looked at longer. After page 11, participants took the first question set (set1) but did not appear to improve their recognition of what the most important areas of content pages are. This behavior has also been mentioned in the previous section. The first question set failed to attract learners' attention to the most important

[2] https://gazeintelligence.com/smi-software-download

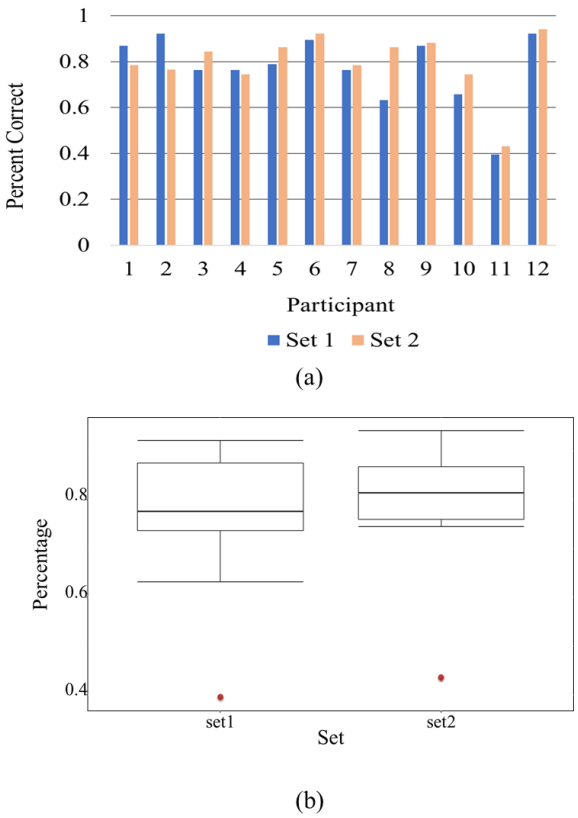


Figure 3: (a) The graph shows that Participants 1, 2, and 4 had fewer correct answers in the second set of questions than the first. All other participants increased their grade. (b) The boxplot for the percentage of correct answers for the question sets 1 and 2. The median for the first set is around 77% with more spread. The median for the second set is approximately 82%. Both question sets have an outlier with a much lower grade (P11). Therefore, they are consistent with the overall trend of improved performance over time.

content of the learning material that we wanted learners to learn and did not alter the learners' visual search pattern.

Figure 6 shows that participants looked at AOI 4 more than the salient AOIs 1-3. Therefore, they spent a lot of time on material that could not help them to answer the questions.

Inferential analysis

To identify important variables that predict the ability to answer correctly, two analysis techniques were used: the random forest classification and regression based on Breiman (2001). The analysis particularly shows that the Normalized Dwell (shown in Figure 6b) is the most critical variable among the eye-tracking data and has the most contribution.

We used an ANOVA analysis to examine the group difference in AOI for correct answers. The correct answer group contains 942 samples. The one-way ANOVA test result showed significant differences between AOI group means as determined by one-way ANOVA ($F(2, 940) = 83.02$, $p < .001$). The ANOVA results indicate that the mean difference of Normalized Dwell over AOI is statistically

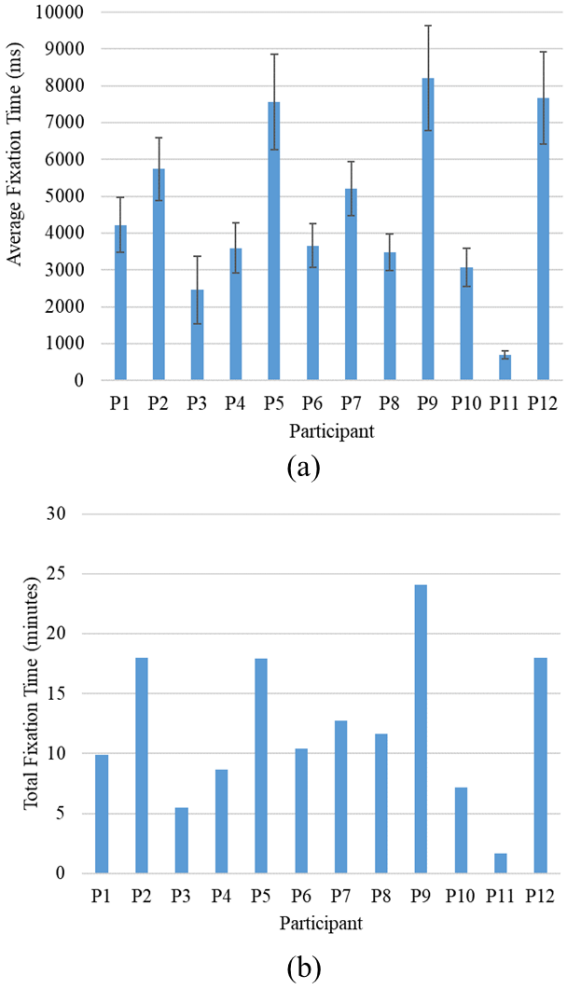


Figure 4: (a) Average fixation time on all AOIs 1-3 on a content page per participant with standard errors shown as error bars. (b) Total fixation time per participant for AOIs 1-3. P11 spent the least amount of time on AOIs 1-3.

different. For instance, AOI 2's mean of Normalized Dwell time is 94,729 ms longer than AOI 1's on average.

The research questions lead to several further analysis questions. To answer research questions, a mixed-effects logistic regression is fitted. The logistic regression was used because the response variable for each question, grade, is binary. Also, a mixed-effects logistic regression was used because the participant is considered to have a random effect in the learner's model. The participant is random because this allows the conclusions made from the model to be extended to others, not just the 12 participants that took part in the study and control the estimation biased. Consequently, what did we find out about the questions noted earlier?

Question 1: Can we predict a learner's performance by using only their visual attention (eye movement data)? Does a model with a learner's eye movement data better predict the learner's performance than a random model? To answer the first question of whether we can predict a learner's performance from their eye movement data, a mixed-effects

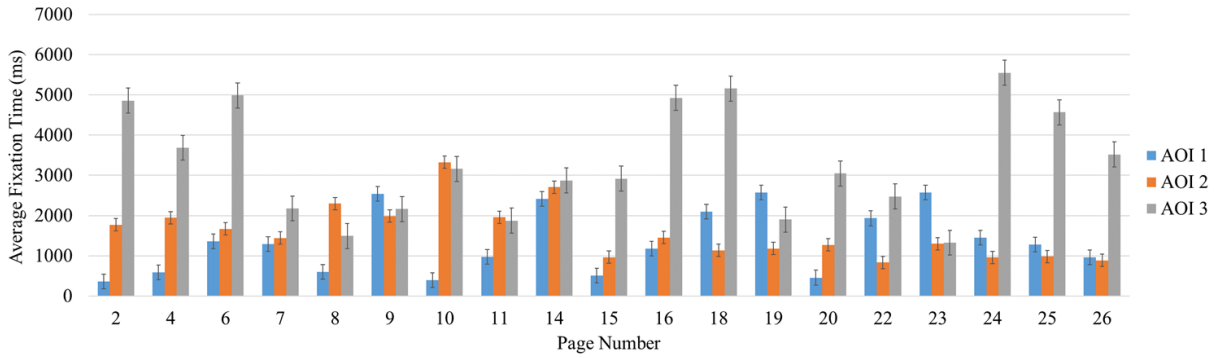


Figure 5**:** Average fixation time for the three primary AOIs per content page. Pages not shown are question pages or non-content pages.

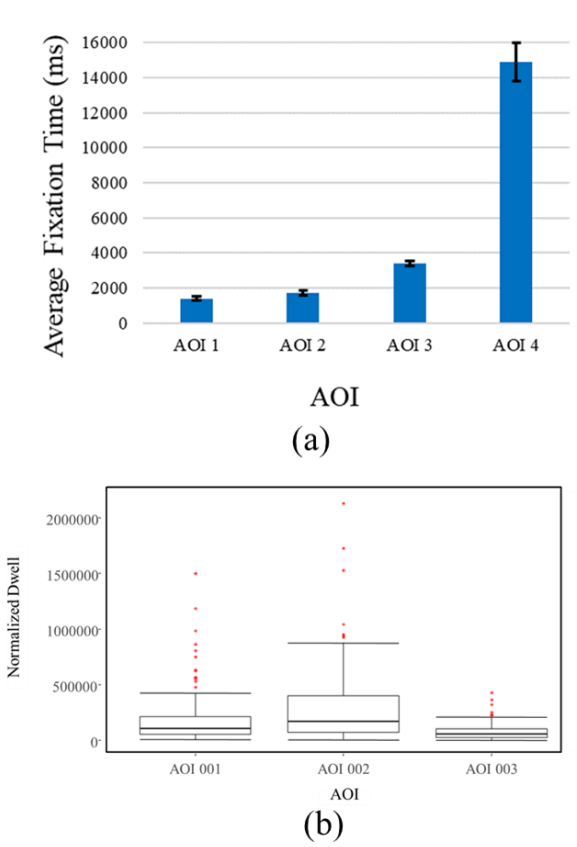


Figure 6: (a) AOIs' average fixation time for all content pages and participants with standard error bars. AOI 4 contains white space, description of GUI items, page numbers, and page content that are not necessary for answering questions. (b) The boxplot for the Normalized Dwell of correct answers for AOIs 1-3.

logistic regression was fitted with only the eye movement data. This method was used to predict grade (on each question), and a Receiver Operating Characteristic (ROC) curve was fit, as well as an AUC value calculated. Figure 7 shows the ROC curve of predicting performance. This curve corresponds to an AUC value of .758, with higher values closer to 1 being a better fit.

Question 2: Can we predict a learner's performance with visual attention and current performance (Set)? To address this question in determining if the variable Set is positively associated with the grade. A mixed-effects logistic regression model was fit with the eye movement variables as well as the Set variable. Multicollinearity assumption was also checked. After the assumptions have been met, the p-value of the Set variable (p=.00095) suggests that the Set variable is a reliable factor in this model. That is, participants perform better in the second question set.

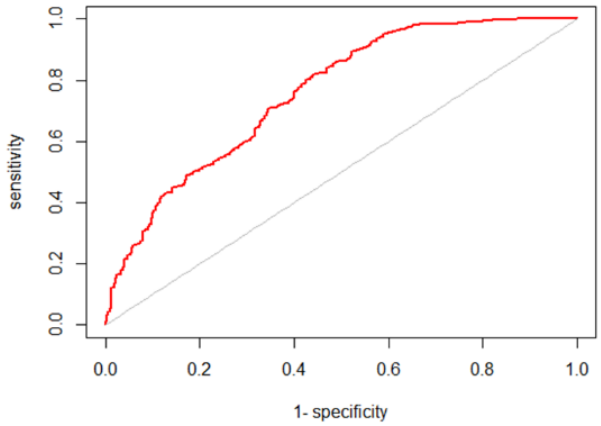


Figure 7: ROC curve of predicting the participant's grade from their eye movement data. 1-specificity is false positive rate and sensitivity is true positive rate.

Question 3: Are the learners' age, gender, where they look, how long they look, the sequence of AOIs, and first language significant predictive factors of the learner's performance? Do they matter when given eye movement data such as fixation time? Are the non-eye movement data variables of Age, Gender, Language, AOI Name, and AOI Coverage significant? To answer these questions, a logistic regression was fit for the non-eye movement variables to predict correct answers. The multicollinearity assumption was checked and

met transfer variables condition—cubic root was used in the equation to transform data.

The results of the logistic regression show that only the variable AOI Coverage was significant. Therefore, the result suggests that the salient areas need to be prominent to attract a participant's attention.

In addition, all data; non-eye movement and eye movement data were considered to find variables that are significant in the model. We again use a mixed-effects logistic regression. The variables that were found significant at the alpha = .05 level were AOI Coverage, Set, the cubed root of Normalized Dwell, and the cubed root of First Fixation. Normalized Dwell and AOI Coverage results suggest that salient areas that take up more space in stimulus have more effect on gaze time and the learner's performance. The first fixation duration suggests that for attracting and maintaining learner's attention the answer's area needs to have a higher priority in the visual scene because only the first visit duration makes a difference, not the total revisiting duration.

Cross-validation of analyses

To analyze what lead to correct answers, we used the KNIME Analytics Platform data mining toolkit[3]. KNIME is an integrated open-source tool that provides a wide choice of advanced data mining algorithms. For the model comparison and model performance, we examined the model using multiple classification algorithms such as Decision Tree, Neural Network, SVM, and Naïve Bayes modules. Previous research (Naik & Samant, 2016) has reported that in the KNIME tool these classification algorithms have higher accuracy compared with other tools. We created scripts to map questions and AOI. Also, we parsed the data and cleaned eye-tracking errors that generate invalid gaze samples. Instead of predicting the overall performance, we used the classification algorithms to predict the individual grade for each question.

Table 1 shows accuracy and kappa (Landis & Koch, 1977) scores for the different classifiers and feature sets for our data and previously published analyses. The kappa value near zero indicates that some additional data may be required for stable and reliable results. Furthermore, because most of the data points in our training data are for correct answers, the expected disagreement is very low in our study. In our case, using kappa value measurements is not a suitable reliability statistic because we do not have enough data from negative answers.

For comparison, the D2P tutor is more straightforward than the MetaTutor that was used in the Bondareva et al. (2013) study. In their research, learner's eye movements were collected while learners watched video tutorials and used MetaTutor tools. In addition to the mouse device, users could type text to the system that caused noise in the data. The main features were representing general gaze trends.

We did not gather the data related to saccades such as gaze direction and path angle; instead, we recorded the visited AOIs sequence. We have additional features for AOIs and participants. The AOI definition in our tutor is based on question answers; however, in Bondareva et al.'s (2013) work, it is based on the type of content (e.g., AOI for all images). In our study, we considered the participant's first language because, in pilot work, we observed that the eye path of participants appears to be affected by their first language.

Table 1: The four data mining analyses were compared with the previous works.

Training Algorithm	Accuracy	Kappa
Decision Tree	77.47%	.39
Neural Network	79.05%	.41
SVM	76.28%	.17
Naïve Bayes	69.96%	.25
Naïve Bayes AOI-only features (Bondareva et al., 2013)	69.6%	.39
Simple Logistic Regression on full dataset (Bondareva et al., 2013)	78.3%	.56

Discussion and conclusion

We analyzed how learners used a tutor by observing where and how long they looked at the information. Their learning was assessed with quiz questions. We found that more time gazing at relevant materials led to more correct answers to the quiz questions. We also saw that for all learners, a significant amount of time was spent gazing at irrelevant material. This effect should be examined to see if these areas can be removed or made less attractive to learners.

We are able to predict learners' performance with only their eye movement data. The high AUC value of .758 in this context means the model is able to correctly predict a learner answering the question correctly about 76% of the time. With the model fit, we are not necessarily able to say we can predict a learner's performance on the second set of questions based on only the first set. However, the mixed-effects logistic regression model showed that the (question) Set variable is significant with a positive coefficient, which means that the participants are more likely to answer a question correctly in the second set of questions than they are in the first. When looking at just the non-eye movement data, only AOI Coverage is significant. Age, gender, and first language did not affect the participant's performance. The random forest analysis, of the eye-movement data, shows the Normalized Dwell's effectiveness in the model. The outcome results illustrate that personal characteristics such as age are not significant, but the Normalized Dwell is a critical factor in the learner's models.

These results suggest that instructional designers and tutors may be able to use this prediction as part of a user model with acceptable accuracy in the early stages of the interaction to

[3] https://www.knime.com/

make adaptive tutors. They can use the relation between the answer's area size and the learners' attention. We have provided several design suggestions for defining answer's areas to have more productive learning experiences.

There are several limitations to this work. The subjects we used were not necessarily typical learners; they were researchers of learning themselves, many with advanced degrees or working toward them. This limits the generality of these findings. Only one tutor was used and for a limited time. Finally, the data set could have had more incorrect answers, which, in turn, would improve the prediction.

There are several ways this work can be extended. We would like to examine further subjects with a broader variety of tutors. We also want to explore the use of these measures of gaze (and perhaps time on a page) to adjust how quickly a tutor progresses a learner through the material.

Our goal was to understand more about what eye tracking can provide in designing an online tutoring environment and intelligent tutors. In this study, we showed that eye tracking can help researchers understand the complete user experience and engagement and make concrete suggestions about how to improve interfaces for learning.

Acknowledgments

This work has been supported by ACT, and ONR grant N00014-15-1-2275. We want to thank Jay Thomas, who provided useful comments on eye-tracking technology. Also, Steve Polyak and Kurt Peterschmidt in ACT Inc. supported us in this study.

References

Anderson, J. R. (2007). *How can the human mind exist in the physical universe?* New York, NY: Oxford University Press.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes. *Psychological Science*, *15*(4), 225-231.

Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J. M., Azevedo, R., & Bouchet, F. (2013). Inferring learning from gaze data during interaction with an environment to support self-regulated learning.In *International Conference on Artificial Intelligence in Education* (pp. 229-238).

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Clinton, V., Cooper, J. L., Michaelis, J. E., Alibali, M. W., & Nathan, M. J. (2017). How revisions to mathematical visuals affect cognition: evidence from eye tracking. In *Eye-Tracking Technology Applications in Educational Research* (pp. 195-218): IGI Global.

Faraday, P. (2000). Visually critiquing web pages.In *Multimedia'99* (pp. 155-166).

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

Holzinger, A., Kickmeier-Rust, M. D., Wassertheurer, S., & Hessinger, M. (2009). Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. *Computers & Education*, *52*(2), 292-301.

Kardan, S., & Conati, C. (2012). Exploring gaze data for determining user learning with an interactive simulation.In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 126-138).

Kardan, S., & Conati, C. (2013). Comparing and combining eye gaze and interface actions for determining user learning with an interactive simulation.In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 215-227).

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., . . . Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, *10*, 90-115.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Lin, Y. T., Wu, C. C., Hou, T. Y., Lin, Y. C., Yang, F. Y., & Chang, C. H. (2016). Tracking students' cognitive processes during program debugging—An eye-movement approach. *IEEE transactions on education*, *59*(3), 175-186.

Naik, A., & Samant, L. (2016). Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, *85*, 662-668.

Ritter, F. E., Yeh, K.-C., Cohen, M. A., Weyhrauch, P., Kim, J. W., & Hobbs, J. N. (2013). Declarative to procedural tutors: A family of cognitive architecture-based tutors.In *Proceedings of the 22nd Conference on Behavior Representation in Modeling and Simulation* (pp. 108-113). Centerville, OH.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols.In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). ACM Press.

Worsley, M., Barel, D., Davison, L., Large, T., & Mwiti, T. (2018). Multimodal interfaces for inclusive learning.In *International Conference on Artificial Intelligence in Education* (pp. 389-393).