Wong, L.-H. et al. (Eds.) (2013). Proceedings of the 21st International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

Understanding Student Interactions with Tutorial Dialogues in EER-Tutor

Myse ELMADANI*, Antonija MITROVIC & Amali WEERASINGHE
Intelligent Computer Tutoring Group, University of Canterbury, New Zealand
*myse.elmadani@pg.canterbury.ac.nz

Abstract: Eye-movement tracking is a potential source of real-time adaptation in a learning environment. In order to have a more comprehensive and accurate picture of a user's interactions with a learning environment, we need to know which interface features he/she visually inspected, what strategies they used and what cognitive efforts they made to complete tasks. Such knowledge allows intelligent systems to be proactive, rather than reactive, to users' actions. Tutorial dialogues is one of the strategies used by Intelligent Tutoring Systems (ITSs) and has been empirically shown to significantly improve learning. EER-Tutor is a constraint-based ITS used to teach conceptual database design. This paper presents the preliminary results of a project that investigates how students interact with the tutorial dialogues in EER-Tutor using both eye-gaze data and student-system interaction logs. Our findings indicate that advanced students are selective of the interface areas they visually focus on whereas novices waste time by paying attention to interface areas that are inappropriate for the task at hand. Novices are also unaware that they require help with the tutorial dialogues.

Keywords: Eye-tracking, tutorial dialogues, constraint-based intelligent tutoring system

1. Introduction

Tutorial dialogues are used in ITSs to simulate one-on-one dialogues between a student and a human tutor that would occur in a classroom environment. The use of tutorial dialogues has been empirically evaluated and shown to significantly improve learning (Olney, Graesser, & Person, 2010; Weerasinghe, Mitrovic, Thomson, Mogin, & Martin, 2011).

This paper outlines work in progress that investigates how students interact with tutorial dialogues using both eye gaze data and student-system interaction logs. From observation, students interact differently with tutorial dialogues in an ITS. One of the obvious factors behind this is the amount of prior knowledge. We believe that we can detect sub-optimal student behaviour from eye-tracking and/or ITS logs in order to allow an ITS to intervene when needed and better guide students' learning. For example, we can investigate if the dialogue contents are being visually attended.

In order to have a more complete picture of a user's interactions with a learning environment, we need to know which interface features he/she visually inspected, the strategies used and what cognitive efforts he/she made to complete tasks (Bednarik, 2005). Such knowledge allows intelligent systems to be proactive, rather than reactive, to users' actions (Gertner & VanLehn, 2000). Some of this information can be obtained by analysing existing student-system interaction logs but Bednarik (2005) highlighted the potential for using eye-movement tracking as a source of real-time adaptation.

Tutorial dialogues have been used in a number of ITSs in order to encourage students to self-explain. Self-explanation is a constructive activity during which a person tries to make sense of new information by explaining it to his or herself (Chi, 2000). This results in the revision of his/her knowledge structure for future application. Despite the effectiveness of ITSs, some studies indicate that students only gain shallow knowledge that they then have difficulty applying to new and different problems (Aleven, Koedinger, & Cross, 1999). One of the ways to overcome this shallow learning problem is to encourage self-explanation in order to promote deep learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Why2-Atlas (Vanlehn et al., 2002) and Auto Tutor (Graesser et al., 2003) teach mainly through tutorial dialogues. In contrast, systems like Geometry Explanation Tutor (Aleven, Ogan, Popescu, Torrey, & Koedinger, 2004) and KERMIT-SE (Weerasinge & Mitrovic,

2005) support problem solving, and tutorial dialogues are provided as additional support. The Geometry Explanation Tutor allows students to give natural language explanations about their problem-solving steps. KERMIT-SE asks students to justify problem-solving decisions that led to an incorrect solution.

The project will give us a better understanding of how different students interact with tutorial dialogues in an ITS. We will get an indication about whether eye-tracking gives us a more complete picture of students' interactions and whether the cost of eye-tracking can be justified. For example, we may be able to use time-based evidence to determine if a student is not paying much attention to the tutorial dialogue by looking at the time between the dialogue appearing and the student responding. This does not reveal situations in which the student is taking time to respond but is visually attending to irrelevant areas of the interface however. A comparison of the accuracy of classifying students' behaviour based on these different measures would be useful. Because we have already implemented adaptive tutorial dialogues in EER-Tutor (Weerasinghe et al., 2011), we will use this constraint-based ITS. EER-Tutor teaches database design using Enhanced Entity-Relationship (EER) data modelling (Elmasri & Navathe, 2007).

We present EER-Tutor in the following section, and discuss related work on using eye-tracking data in Section 3. Section 4 presents the experiment we performed, followed by a discussion of results in Section 5. Finally, we present conclusions and future research plans in Section 6.

2. EER-Tutor

EER-Tutor (Zakharov, Mitrovic, & Ohlsson, 2005) is a constraint-based ITS which provides a learning environment for students to practise and learn database design using the EER data model. The interface of EER-Tutor (Figure 1) shows the problem statement at the top, the toolbox containing the components of the EER model, the drawing area on the left, and the feedback area on the right (please note that the screenshot in Figure 1 shows the version of EER-Tutor used in the study, not the standard interface). Students create EER diagrams satisfying a given set of requirements which are checked for constraint violations on submission. EER-Tutor records detailed session information, including each student's attempt at each problem and its outcome.

Tutorial dialogues have been implemented for EER-Tutor (Weerasinghe, Mitrovic, & Martin, 2008; Weerasinghe et al., 2011). The model for supporting adaptive dialogues is out of the scope of the current paper, and we refer the interested reader to (Weerasinghe, Mitrovic, & Martin, 2009) for details. Here we present only a short explanation of the tutorial dialogues.

When a student makes one or more mistakes, he/she is presented with a tutorial dialogue selected adaptively. The problem statement, toolbar and drawing area are disabled but visible for the duration of the dialogue and the error is highlighted in red in the diagram. Each dialogue consists of several prompts, and provides multiple possible answers to the student. Therefore the student answers prompts by selecting an option he/she believes is correct, or asks for additional help by selecting the "*I need more help*" option. The prompt types analysed are as follows:

- **Conceptual (CO):** discusses the domain concept with which the student is having difficulty independently of the problem context. This is shown in Figure 1: the student has modelled 'Address' as a derived attribute instead of a simple attribute so the prompt is asking about the basics of derived attributes. An incorrect answer to a *conceptual* prompt results in an additional, simpler *conceptual* prompt being given.
- **Reflective (RE)**: aims to help students understand why their action is incorrect in the context of the current problem, and therefore the prompt text refers to the elements of the problem. For the error in Figure 1, this is: "Can you tell me why modelling Address as a derived attribute is incorrect?"
- **Corrective action (CA)**: gives the student the opportunity to understand how to correct the error for this specific problem. For the error in Figure 1, the CA prompt is asking the student to specify the best way to model the 'Address' attribute and giving the different attribute types as options. Not all dialogues have this prompt type.

- **Conceptual reinforcement (CR)**: allows the student to review the domain concept learned. For the error in Figure 1, the CR prompt asks the student to choose the correct definition of a derived attribute from the given options. Again, this is a problem-independent prompt.

In addition to multi-level dialogues, students are given single-level hints when they make basic syntax errors such as leaving diagram elements unconnected. When we use prompts and dialogues from now on we only refer to multi-level dialogues unless otherwise specified.

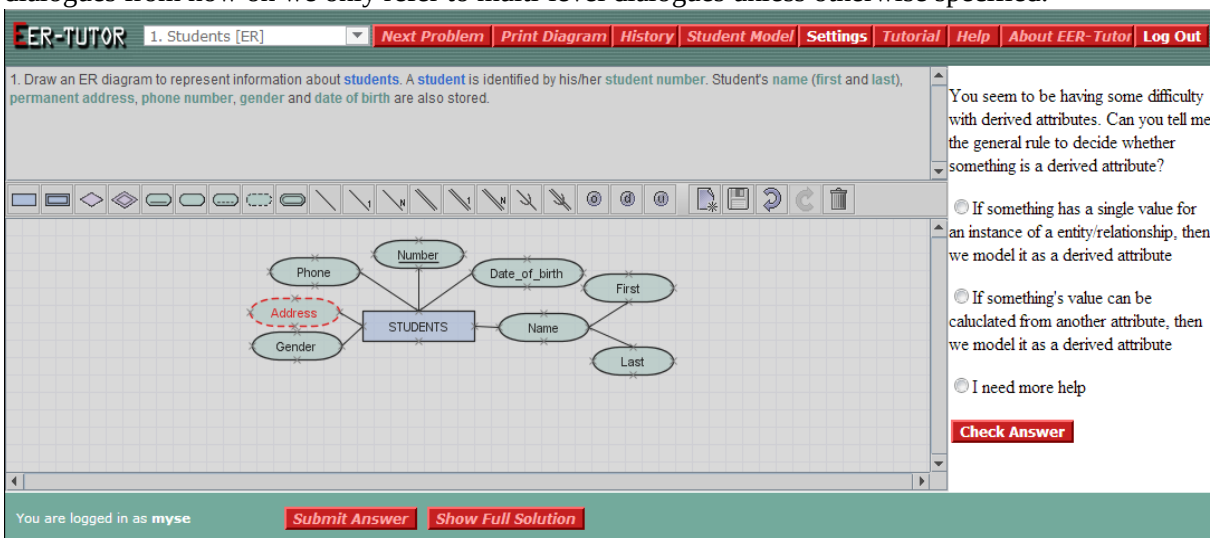

Figure 7. The EER-Tutor interface

3. Related Eye-tracking Research

Eye-tracking is concerned with determining the point of gaze of a person's eyes. The human eye makes a series of rapid eye movements (saccades) then fixates on some point in a visual scene (a fixation) (Goldberg & Helfman, 2010). In order to help with eye-tracking analysis, Areas of Interest (AOIs) can be set up that specify important regions in the user interface. AOIs help by defining which fixations should be tallied and to define scanning sequences and transitions for example (Goldberg & Helfman, 2010). Eye-tracking is used in interface usability studies, advertising as well as developmental psychology. In regards to ITSs, eye-tracking has been used in a variety of research, ranging from error prediction and determining if students read system feedback (Gluck, Anderson, & Douglass, 2000), to its use as a form of input (Wang, Chignell, & Ishizuka, 2006) and to the analysis of how students interpret open learning models (Mathews, Mitrovic, Lin, Holland, & Churcher, 2012).

3.1 Eye-tracking for User Modelling

The use of eye-tracking to increase student model bandwidth for educational purposes was first discussed in (Gluck et al., 2000). Students used the EPAL Algebra Tutor, an adaptation of the Worksheet tool in Algebra Tutor (Koedinger & Anderson, 1998). Gluck, Anderson et al. (2000) found that students ignore bug messages and that some students also ignore algebraic expressions so the tutor can draw their attention to these areas.

Conati and Merten carried out online assessment of students' self-explanation using eye-tracking data and interface actions (Conati & Merten, 2007). Students used the Adaptive Coach for Exploration (ACE), an exploration-based learning environment that allows students to explore mathematical functions (Bunt, Conati, Huggett, & Muldner, 2001). An empirical evaluation of a probabilistic user model including self-explanation found that gaze-tracking data improved model performance compared to using only time data as a predictor of self-explanation.

In other work, Kardan and Conati (2012) used only users' visual attention patterns to assess their ability to learn with an interactive simulation. The simulation used shows how an algorithm for solving constraint satisfaction problems in the domain of Artificial Intelligence works. The authors

also found that the changes in users' attention patterns when moving to solving a more difficult problem can be used to classify the students based on performance. For example, high achievers increase the number of fixations on an AOI that should be used more in the harder problem.

Eivazi and Bednarik (2011) proposed the use of real-time tracking of users' visual attention patterns to model users' high-level cognitive states and performance. The rationale is that an intelligent system can monitor users and use eye movement data to guide learning. Eye movement features were calculated for each interval corresponding to an utterance coded to a cognitive trait such as planning while solving an 8-tile puzzle. A Support Vector Machine-based classification was used to predict problem-solving cognition states such as planning as well as a user's performance. Performance was accurately predicted: the high-performance group had a lower number of fixations but longer fixation durations than the low-performance group for example.

Tsai et al. (2012) used eye-tracking to study students' visual attention when predicting debris slide hazards in an image-based, multiple-choice science problem. Students attended more to the chosen option than the rejected ones and spent more time inspecting features relevant to the answer chosen than features that are irrelevant to it. Regarding successful problem solvers, the study found that they shift gaze from irrelevant to relevant features. This is in contrast to unsuccessful problem solvers who shift their gaze from relevant to irrelevant features and the problem statement.

3.2 Eye-tracking and Expertise

Differences between novice and expert pilots were also found using eye-tracking data gathered during simulated Visual Flight Rules flight landings (Kasarskis, Stehwien, Hickox, Aretz, & Chris Wickens, 2001). Experts had more frequent fixations on relevant areas but for shorter durations. The scan patterns of experts are also stronger and more defined, which means that they better maintain airspeed and good landing performance because the patterns are consistent and efficient.

Law et al. (2004) also used of eye gaze patterns to differentiate between novices and experts. Novice and expert surgeons performed a task on a computer-based laparoscopic surgery simulator. Expert surgeons were quicker and made fewer errors overall as expected. Novices had to fixate on the tool's position and had varied behaviours whereas experts could manipulate the tool while maintaining eye gaze on the target.

Jarodzka et al. (2010) similarly investigated the differences in strategies used by novices and experts and therefore the areas they fixate on. When asked to describe the locomotion patterns of swimming fish from a video, experts attended to task-relevant features more than novices and remained focussed on these areas. In addition, experts focused on different features because they employed knowledge-based shortcuts unknown to novices.

4. The Study

During the week prior to our study, the participants had used EER-Tutor in a regular lab session and completed a pre-test. This test was made up of 6 questions and included these types of questions: problem solving (drawing an EER diagram for the given scenario), multiple-choice, short answer and justification. The maximum mark on the pre-test was 27.

27 Computer Science students (9 females), aged from 18 to 50 years old (mean 23.8 years, standard deviation 7.3 years) participated in the study. All participants had normal or corrected-to-normal vision. The participants were enrolled in a second-year database course at the University of Canterbury and volunteered to take part in the study. Each participant was given a NZ$20 voucher on completion of the study and took part in the study individually.

The EER-Tutor version used in the study excluded interface features unneeded for the study (scrolling for example). The dialogue prompts and options vary in length so we always displayed the options in the same position to ease the definition of areas of interest. In addition, the tutorial dialogues were not adaptive. That is, all students received the same order of prompts within a dialogue about the same error (constraint violation). This means that the dialogue length is only affected by the correctness of the students' answers.

We used the Tobii TX300 Eye Tracker, which allows unobtrusive eye-tracking as it is an integrated eye-tracker. Subjects are able to move during the tracking session while accuracy and precision are maintained at a sampling rate of 300 Hz (Tobii Technology AB, 2010).

The participants initially read an information sheet and signed a consent form. They then provided their age and vision status. A calibration phase with the eye-tracker was then carried out. This involves the participant following a marker on a 9-point grid with their eyes. The participants were instructed to complete or at least attempt all of the problems and to submit their solutions regularly.

During the session, the students could work on three problems and were free to move between problems. The students build up the diagrams incrementally and are free to choose the order in which they model elements and the elements' positions in the solution area. Each diagram is therefore different, which differs from the majority of the related work above. Two problems were of moderate difficulty, and the last one was the most difficult. We selected problems that describe real-world situations the participants were familiar with (a health club, student accommodation services and the Olympic games), but the problems are not too easy so that students make mistakes when working on them and get tutorial dialogues corresponding to the mistakes made. Each student was given 50 minutes to solve the problems. Participants were reminded to regularly submit their solutions during the session. The mean session length was 49.1 minutes (standard deviation 3.0 minutes).

5. Data Analyses

One participant was excluded because no eye-tracking data was collected. We classified the participants as *Novice* or *Advanced* using a median split on pre-test scores (median score is 13.50). This resulted in 13 novices (Mean=10.50, SD=2.30) and 13 advanced students (Mean=16.50, SD=1.94). Using a median split means calculating the median pre-test score for the participants and using this as the threshold for defining groups: a novice is a student with a pre-test score of 13.50 or less and an advanced student has a pre-test score greater than 13.50. Because of the small sample size, we used non-parametric statistical analysis methods. A Mann-Whitney U-Test shows that there is a significant difference in the distributions of pre-test scores between the two groups ($U=0$, $p < 0.001$).

Table 5: Summary statistics of *Novice* and *Advanced* students

	Novice	Advanced	U (Sig.)
Mean number of completed problems (SD)	1.08 (1.12)	2.15 (0.69)	39.50 (**0.019****)
Mean number of submissions per completed problem (SD)	10.29 (5.06)	11.31 (5.17)	39.50 (**NS**)
Mean time per completed problem in seconds (SD)	948.05 (582.66)	1000.13 (636.32)	37.00 (**NS**)
Mean time per attempted problem in seconds (SD)	1212.73 (238.44)	1042.69 (223.48)	127.00 (**0.029****)
Mean number of submissions (SD)	19.15 (8.32)	19.46 (8.26)	77.00 (**NS**)
Mean number of single-level dialogues seen (SD)	0.91 (0.83)	0.74 (0.49)	84.50 (**NS**)
Mean number of dialogues seen (SD)	15.77 (6.70)	15.08 (8.09)	86.50 (**NS**)
Mean number of prompts seen (SD)	53.92 (22.30)	44.92 (26.26)	106.00 (**NS**)
Mean dialogue length (SD)	3.43 (0.21)	2.99 (0.45)	158.00 (<**0.001****)
Mean time per prompt in seconds (SD)	17.03 (7.94)	15.78 (6.78)	93.00 (**NS**)
Mean number of unique relevant constraints (SD)	113.77 (8.88)	117.69 (3.28)	70.50 (**NS**)
Mean number of violated constraints (SD)	128.08 (44.51)	105.23 (83.90)	124.50 (**0.039****)

5.1 Analyzing the EER-Tutor Logs

There were 502 submissions (i.e. solution attempts) in total and 1285 prompts seen. Table 1 shows a summary of the statistics for the *Novice* and *Advanced* student groups. The distribution of each statistic across groups was tested using the Independent-Samples Mann-Whitney U test with $\propto$=0.05. This test is used for all *Novice-Advanced* comparisons in this paper.

Table 6: A comparison of the percentage of choices made by *Novice* and *Advanced* students

	Novice	Advanced	U (Sig.)
Mean percentage of correct choices (SD)	78.13 (13.46)	85.36 (9.34)	59.00 (**NS**)
Mean percentage of help choices (SD)	2.63 (2.93)	1.95 (3.19)	104.00 (**NS**)
Mean percentage of incorrect choices (SD)	17.62 (10.75)	10.88 (8.26)	117.5 (**0.10***)
Mean percentage of unanswered prompts (SD)	1.62 (1.55)	1.81 (2.31)	84.00 (**NS**)
Mean percentage of correct choices for reflective prompts (SD)	73.11 (16.01)	84.71 (12.59)	49.00 (**0.072***)
Mean percentage of incorrect choices for reflective prompts (SD)	21.60 (14.10)	9.08 (8.84)	133.00 (**0.012****)

As expected, the distributions of the mean number of completed problems are significantly different. *Advanced* students solved more problems on average but the distributions of the mean number of submissions and time spent per completed problem are not significantly different. When we also consider attempted problems that were not completed, we see that there is a significant difference in the distributions of the mean time spent per problem. In fact all of the *Novices* attempted the problems in order of difficulty and only 6 attempted the most difficult problem (compared to 12 *Advanced* students).

Because the number of submissions, dialogues (both single- and multi-level) and prompts seen are not significantly different, we analysed finer-grained measures. The distributions of the

dialogue length are significantly different. We expected this result as *Novice*s may not always answer prompts correctly because they have misconceptions or missing domain knowledge (reflected by dialogue length). Interestingly the distributions of the time per prompt are not significantly different.

The distributions of the number of unique relevant constraints are not significantly different because all students were solving the same problem set. The distributions of the mean number of violated constraints are significantly different as expected however.

A comparison between the number of prompts of each type seen by *Novice* and *Advanced* students did not reveal any unexpected results. Regarding the correctness of the choices made for each prompt, it was expected that *Advanced* students would make fewer *incorrect* choices because they began with more prior domain knowledge. The distributions are marginally different as seen in Table 2. Although the distributions are not significantly different, it is interesting that *Advanced* students make more *correct* choices on average as expected. We had expected there to be some difference in the percentage of *help* choices made by the two groups however, suggesting that *Novices* may be unaware that they require assistance in answering the prompts.

Because we are interested in students' interactions with the tutorial dialogues, it is useful to look at the time spent and prompt answer correctness for each prompt type. We had expected to find some differences in the distributions of mean prompt reflection times for the different prompt types but this was not the case. While a conceptual prompt may not require the student to read the problem statement or spend time reflecting on their solution, this is expected behaviour when interacting with a reflective prompt for example. The distributions of *correct* and *incorrect* answers for *reflective* prompts are marginally and significantly different respectively (see Table 2). These findings are not surprising but it is interesting that this is the only prompt type with differences between the two groups.

5.2 Analyzing the Eye-Tracking Data

For the data during which prompts were visible, we output the following metrics from Tobii Studio:

- **Fixation duration (seconds):** duration of each individual fixation in an AOI.
- **Fixation count:** number of times the participant fixates on an AOI.
- **Visit count:** number of visits to an AOI.

Table 7: Summary eye-gaze metrics comparing *Novice* and *Advanced* students

	Novice	**Advanced**	**U (Sig.)**
Mean fixation duration (SD)	0.26 (0.06)	0.23 (0.04)	114.50 (**NS**)
Mean fixation count (SD)	226.23 (107.09)	135.01 (68.07)	131.00 (**0.016****)

Table 3 shows the above metrics for *Novice* and *Advanced* students for the whole EER-Tutor interface (visit count is excluded as there is only one AOI). The distributions of the mean fixation duration are not significantly different, with a shorter mean fixation duration for *Advanced* students. More informative results should be possible when we break down the EER-Tutor interface into key areas especially because the distributions of the mean fixation count are significantly different. *Advanced* students are making fewer fixations and so it would be useful to see the differences between where *Novices* and *Advanced* students look.

The AOIs defined for the EER-Tutor interface are shown in Figure 2. The *problem* and *feedback* are where the problem text and dialogues are displayed respectively. Users build their diagram on the *canvas,* and the *toolbar* has been included because it would be interesting to see if it is being used despite it being disabled during dialogues. The length of the prompt text and options varies so extra space has been included to ensure they are displayed in the same position in order to make it possible to analyse finer-grained metrics for the *feedback* AOI (for example transitions between the three options).

Another reason to use AOIs is to make the eye-gaze patterns clearer by defining regions with specific functions. For example Figure 3 shows two gaze patterns of the same *Advanced* student: the first pattern is for all *corrective action* prompts and the second for *conceptual* prompts. It is clear that the student is re-reading the problem statement for the *corrective action* prompts but not the

conceptual prompts. This is not surprising because *conceptual* prompts discuss problem-independent domain knowledge while the *corrective action* prompt refers to the error in the solution of a specific problem.

We have not completed any complex gaze-pattern analysis to date so Table 4 shows the analysis of the above eye-gaze metrics but breaks down the results by both the prompt type and AOI. Only marginally and significantly different results are included.

The distributions of the mean fixation durations for the *canvas* AOI are different for both *conceptual* and *conceptual reinforcement* prompts. For *conceptual reinforcement* prompts there is also a significant difference in the distributions of the mean fixation durations for the *problem* AOI. This is interesting because these prompt types are problem-independent and so there is no need to look at the *canvas* or *problem* statement to answer the prompts. The reason for both groups looking at the *canvas* may be the fact that the error is highlighted in red on the diagram so is eye-catching. The difference may be that *Advanced* students do not look at their solutions for as long because they do not need it to answer the prompt. This is supported by the *Advanced* students' lower visit counts to the *canvas* AOI but further investigation is needed.

The distributions of the mean fixation counts are significantly different for the *canvas* and *feedback* AOIs of *conceptual* prompts. The *canvas* and *problem* AOIs' distributions are also significantly different for *conceptual reinforcement* prompts. Again there is no reason to look at the *canvas* other than to look at the highlighted error as discussed above. While *Advanced* students have a lower mean fixation count on the *feedback* area, it may be that they are able to select the correct option more quickly than *Novices*. More details of the students' behaviour would be revealed by breaking down the *feedback* AOI into *prompt text* and individual *option* AOIs.

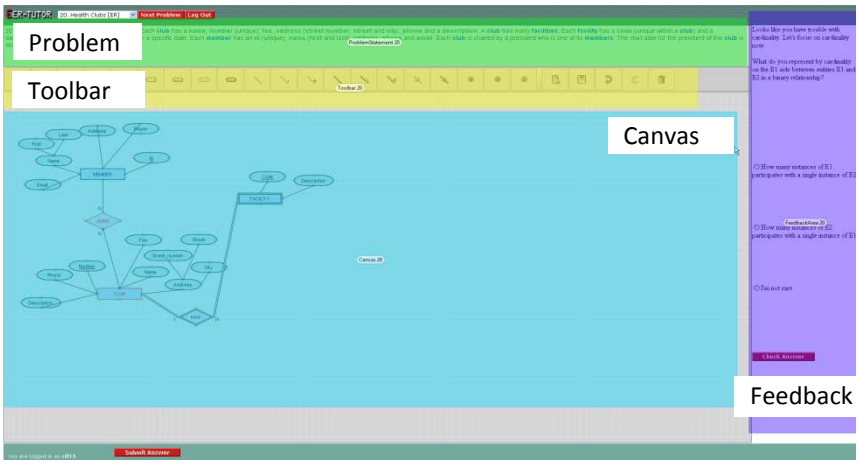

Figure 8: The AOIs defined for the EER-Tutor interface

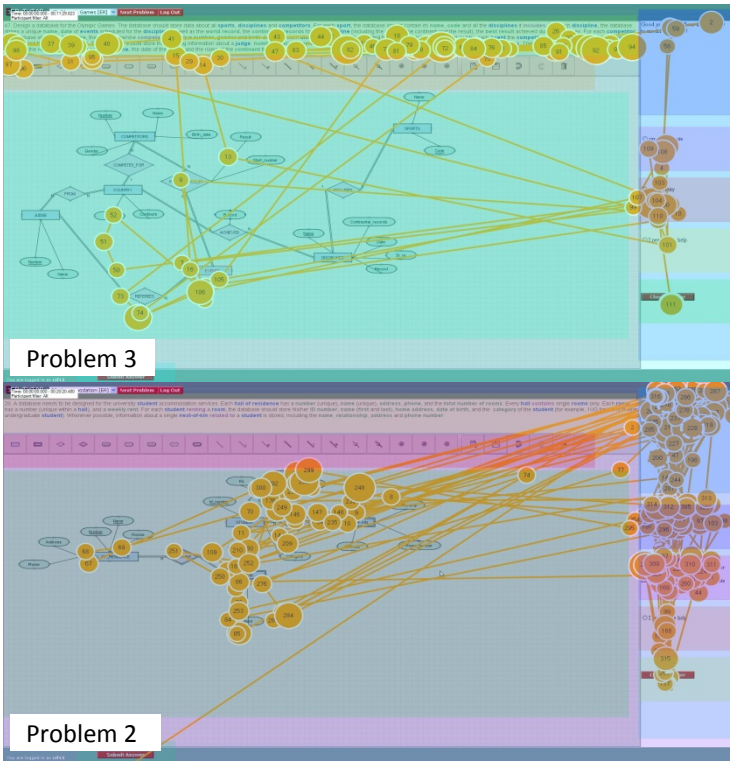

Figure 9: Gaze patterns for one student: *corrective action* prompts (left) and *conceptual* prompts (right)

The distributions of the mean visit counts are marginally different for the *feedback* AOI for *corrective action*, *conceptual reinforcement* and *conceptual* prompts. This can be a result of *Advanced* students making fewer transitions between the AOIs because they recognise the relevant AOIs for each prompt type and focus on those (in this case the *feedback* area) whereas *Novices* are unaware of the relevance of the different areas and therefore more frequently switched between AOIs. In fact when looking at the same *conceptual* prompt, for the same problem for an *Advanced* and a *Novice* student shows that the overall pattern may appear similar initially (see Figure 4). After considering the order of the fixations however, it becomes clear that the *Advanced* student looks at the *canvas* only once before returning and focussing on the *feedback* AOI for the rest of the time the prompt is displayed. This is in contrast to the *Novice*, who switches between the *feedback* and *canvas* more frequently and even looks at the problem statement briefly. Further investigation of the transitions is required however as this is a quick observation of behaviour of these two specific students.

Table 8: Summary eye-gaze metrics comparing *Novice* and *Advanced* students: prompt types and AOIs

	Prompt	**AOI**	**Novice**	**Advanced**	**U (Sig.)**
Mean fixation duration (SD)	CO	Canvas	0.24 (0.02)	0.20 (0.03)	134.50 (**0.001****)
	CR	Canvas	0.22 (0.06)	0.17 (0.08)	112.50 (**0.06***)
	CR	Problem	0.30 (0.09)	0.19 (0.05)	32.00 (**0.026****)
Mean fixation count (SD)	CO	Canvas	70.93 (38.99)	24.32 (13.07)	141.00 (<**0.001****)
	CO	Feedback	214.70 (103.18)	117.44 (87.39)	121.00 (**0.019****)
	CR	Canvas	21.46 (19.73)	8.64 (11.54)	122.00 (**0.016****)
	CR	Problem	8.92 (6.09)	3.08 (1.96)	29.00 (**0.093***)
Mean visit count (SD)	CO	Canvas	17.45 (9.06)	7.49 (5.18)	138.00 (**0.001****)
	CO	Feedback	30.32 (21.01)	17.18 (17.65)	113.50 (**0.052***)
	CA	Feedback	5.96 (3.12)	4.08 (1.70)	86.50 (**0.08***)
	CR	Canvas	5.73 (4.07)	3.91 (4.23)	110.50 (**0.077***)
	CR	Feedback	18.61 (16.95)	11.45(7.94)	90.00 (**0.052***)

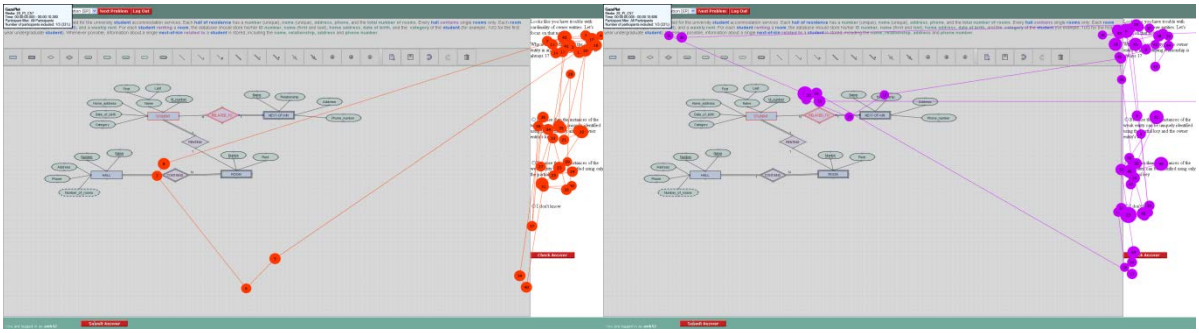

Figure 10: The gaze patterns of the same prompt for an *Advanced* (left) and *Novice* (right) student

6. Conclusion and Future Work

It is evident that there are some differences between *Novice* and *Advanced* students in terms of their behaviour as indicated by the collected EER-Tutor and eye-tracking data. From the EER-Tutor logs we see that there is no significant difference in the distributions of the percentage of *help* choices made by the two groups. *Novices* therefore may need the ITS to intervene and explain the error being discussed in more detail. The distributions of the time-based evidence 'mean time spent per prompt' are not statistically significant however, suggesting a knowledge gap because we are using only EER-Tutor logs. Another data source such as eye-gaze data can be combined with the EER-Tutor log data so that we better understand students. The eye-gaze data reveal that there may be differences between the groups in terms of *Advanced* students being more selective about the areas they focus on and therefore making fewer visits to irrelevant AOIs. While this needs to be investigated further, it demonstrates that it should be possible for an ITS to eventually detect sub-optimal behaviours that produce these effects from both sources in real-time and react appropriately. An example of sub-optimal behaviour is a student who does not refer to his/her solution at all when interacting with the tutorial dialogues.

We intend to further analyse the collected data, in particular by investigating the transitions between the different AOIs of the EER-Tutor interface (the students' gaze patterns) and using machine learning to discover commonly occurring learning behaviours. One of the next steps is to build classifiers that are able to automatically place students into the appropriate group. Classifiers built using only EER-Tutor data, only eye-gaze data and both data sources will be compared. The performance of these classifiers will help us determine whether the cost of eye-tracking is justified: features calculated from EER-Tutor may provide reasonably accurate student classification that is relatively inexpensive to collect. In addition, this work needs to be incorporated into EER-Tutor to provide adaptive interventions and guidance, which themselves are directions for future research.

References

Aleven, V., Koedinger, K. R., & Cross, K. (1999). Tutoring Answer Explanation Fosters Learning with Understanding. In S. P. Lajoie & M. Vivet (Eds.), *Proceedings of the 9th International Conference on Artificial Intelligence in Education, AIED '99* (pp. 199–206).

Aleven, V., Ogan, A., Popescu, O., Torrey, C., & Koedinger, K. (2004). Evaluating the Effectiveness of a Tutorial Dialogue System for Self-Explanation. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Proceedings of Seventh International Conference on Intelligent Tutoring Systems* (pp. 443–454).

Bednarik, R. (2005). Potentials of eye-movement tracking in adaptive systems. In *Proceedings of the Fourth Workshop on the Evaluation of Adaptive Systems* (pp. 1–8).

Bunt, A., Conati, C., Huggett, M., & Muldner, K. (2001). On improving the effectiveness of open learning environments through tailored support for exploration. In *Proceedings of AIED 2001, 10th World Conference of Artificial Intelligence and Education* (pp. 365–376).

Chi, M. T. H. (2000). Self-Explaining Expository Texts: The Dual Process of Generating Inferences and Repairing Metal Models. *Advances in Intstructional Psychology*, *5*, 161–238.

Chi, M. T. H., Bassok, M. M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, *13*, 145–182.

Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems*, *20*(6), 557–574.

Eivazi, S., & Bednarik, R. (2011). Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data. In *2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction* (pp. 9–16).

Elmasri, R., & Navathe, S. B. (2007). *Fundamentals of Database Systems*.

Gertner, A., & VanLehn, K. (2000). Andes: A coached problem solving environment for physics. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent Tutoring Systems: 5th International Conference* (pp. 133–142).

Gluck, K. A., Anderson, J. R., & Douglass, S. A. (2000). Broader Bandwidth in Student Modeling: What if ITS Were " Eye " TS? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *ITS '00: Proceedings of the 5th International Conference on Intelligent Tutoring Systems* (pp. 504–513).

Goldberg, J. H., & Helfman, J. I. (2010). Comparing Information Graphics: a Critical Look at Eye Tracking. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization* (pp. 71–78).

Graesser, A. C., Jackson, G. T., Mathews, E. C., Mitchell, H. H., Olney, A., Ventura, M., … Group, T. R. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty- Fifth Annual Conference of the Cognitive Science Society* (pp. 474–479).

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*(2), 146–154. doi:10.1016/j.learninstruc.2009.02.019

Kardan, S., & Conati, C. (2012). Exploring Gaze Data for Determining User Learning with an Interactive Simulation. In *User Modeling, Adaptation, and Personalization* (pp. 126–138).

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Chris Wickens. (2001). Comparison of expert and novice scan behaviors during VFR flight. In R. Jensen (Ed.), *Proceedings of the 11th International Symposium on Aviation Psychology* (pp. 1–6).

Koedinger, K. R., & Anderson, J. R. (1998). Illustrating principled design: The early evolution of a cognitive tutor for algebra symbolization. *Interactive Learning Environments*, *5*, 161–179.

Law, B., Atkins, M. S., Kirkpatrick, A. E., Lomax, A. J., & Mackenzie, C. L. (2004). Eye Gaze Patterns Differentiate Novice and Experts in a Virtual Laparoscopic Surgery Training Environment. In A. T. Duchowski & R. Vertegaal (Eds.), *Proceedings of the 2004 symposium on Eye tracking research & applications* (pp. 41–48).

Mathews, M., Mitrovic, A., Lin, B., Holland, J., & Churcher, N. (2012). Do your eyes give it away? Using eye tracking data to understand students' attitudes towards open student model representations. In S. Cerri, W. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (pp. 422–427).

Olney, A. M., Graesser, A. C., & Person, N. K. (2010). Tutorial Dialog in Natural Language. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 181–206). Springer.

Tobii Technology AB. (2010). *Tobii TX300 Eye Tracker Product Description*.

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.

Vanlehn, K., Jordan, P. W., Rosé, C. P., Bhembe, D., Böttner, M., Gaydos, A., … Srivastava, R. (2002). The Architecture of Why2-Atlas: A Coach for Qualitative Physics Essay Writing. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Proceedings of the 6th International Conference on Intelligent Tutoring Systems* (pp. 158–167).

Wang, H., Chignell, M., & Ishizuka, M. (2006). Empathic Tutoring Software Agents Using Real-time Eye Tracking. In *Proceedings of the 2006 symposium on Eye tracking research & applications - ETRA '06* (pp. 73–78). New York, New York, USA: ACM Press.

Weerasinge, A., & Mitrovic, A. (2005). Supporting Deep Learning in an Open-ended Domain. In *Studies In Fuzziness And Soft Computing* (Vol. 179, pp. 105–152).

Weerasinghe, A., Mitrovic, A., & Martin, B. (2008). A Preliminary Study of a General Model for Supporting Tutorial Dialogues. In *16th International Conference on Computers in Education* (pp. 125–132).

Weerasinghe, A., Mitrovic, A., & Martin, B. (2009). Towards Individualized Dialogue Support for Ill-Defined Domains. *International Journal of Artificial Intelligence in Education*, *19*(4), 357–379.

Weerasinghe, A., Mitrovic, A., Thomson, D., Mogin, P., & Martin, B. (2011). Evaluating a General Model of Adaptive Tutorial Dialogues. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of the 15th international conference on Artificial intelligence in education* (pp. 394–402).

Zakharov, K., Mitrovic, A., & Ohlsson, S. (2005). Feedback Micro-engineering in EER-Tutor. In C.-K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Artificial Intelligence in Education* (Vol. 179, pp. 718–725). Amsterdam: IOS Press.