Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

Identifying Learner Differences in Example Processing from Eye Gaze Data

Amir SHAREGHI NAJAR, Antonija MITROVIC[*] & Kourosh NESHATIAN
Department of Computer Science and Software Engineering, University of Canterbury, New Zealand
*tanja.mitrovic@canterbury.ac.nz

Abstract: Learning from worked examples (WE) has been shown to be beneficial for novices. We have previously conducted two studies, comparing learning from examples to tutored problem solving in SQL-Tutor, and Intelligent Tutoring System (ITS). The first study showed that interleaving examples with supported problem solving is an optimal choice compared to using either of those two types of learning in isolation. In the second study, we added an adaptive strategy for selecting WE or problems to be given to the learner, which proved to be superior to the fixed sequence of WE and problems. In this paper, we focus on how students with different levels of knowledge process WEs. Our goal is to identify meaningful differences in example processing that can be used to provide adaptive hints to the learner. In order to comprehend SQL examples, the learner needs to understand the database which is used as the context. We analysed eye movements collected from a quasi-experiment, and found a significant difference in the amount of attention students paid to database schemas.

Keywords: Worked examples, eye tracking, learner differences

1. Introduction

Numerous studies have shown that learning from worked examples is beneficial for novices in comparison to unsupported problem solving, e.g. (Sweller, Ayred & Kalyuga, 2011; Kirschner, Sweller & Clark, 2006; Atkinson et al., 2000; Van Gog, 2011). A worked example (WE) consists of the solution and additional explanations, thus providing knowledge that the learner might lack. Therefore, worked examples allow the learner to focus on important concepts, thus greatly reducing the cognitive load (Sweller, Ayred & Kalyuga, 2011). Recently researchers have started comparing learning from examples to Tutored Problem Solving (TPS) in ITSs (Koedinger & Aleven, 2007; Schwonke et al., 2009), showing that learning from WE reduces learning time. Some studies have found no difference in the amount of learnt knowledge (McLaren & Isotani, 2011) between those two modes of learning. Our previous study (Shareghi Najar & Mitrovic, 2013; 2014) showed that learning from alternating WEs with problems is superior to learning from TPS or WEs only, when the sequence of problems/examples is fixed. In a later study, we compared a fixed sequence of WEs and TPS to an adaptive strategy which decided whether to present an example or a problem based on the student's performance (Shareghi Najar, Mitrovic & McLaren, 2014). The adaptive strategy was superior to the fixed alternating sequence.

In order to further improve our adaptive strategy, we decided to investigate whether there are meaningful differences in example processing between students with different levels of SQL knowledge. We decided to collect eye gaze data, which provides fine-grained information about how learners study worked examples.

The student's interactions with the ITS interface is of high importance, as it can reveal what students pay attention to. Therefore, eye-tracking data can be used to improve student modeling and to provide adaptive support (Kardan & Conati, 2012). There are also studies (Bull, Cooke & Mabbott, 2007; Mathews et al., 2012) that have used eye-tracking data to investigate how students interpret various presentations of open student models. Eye tracking also enables investigation of successful and unsuccessful student behaviour that leads the student to learn or fail to learn. ITSs can classify students as novice or advanced students by matching their behaviour to successful and unsuccessful behaviours. This information allows the ITS to provide adaptive support to the student to avoid unsuccessful behaviour and encourage productive behaviour (Kardan & Conati, 2012).

We start by presenting the version of SQL-Tutor used in our study, followed by the experimental design. Section 4 presents the results of the study. The conclusions and the directions of future work are presented in Section 5.

2. SQL-Tutor

SQL-Tutor (Mitrovic, 1998; 2003) is a constraint-based tutor that teaches Structured Query Language (SQL), the most widely used query language for relational databases. SQL-Tutor complements traditional lectures; it assumes that the student has already acquired some knowledge via lectures and labs, and provides problem-solving opportunities.

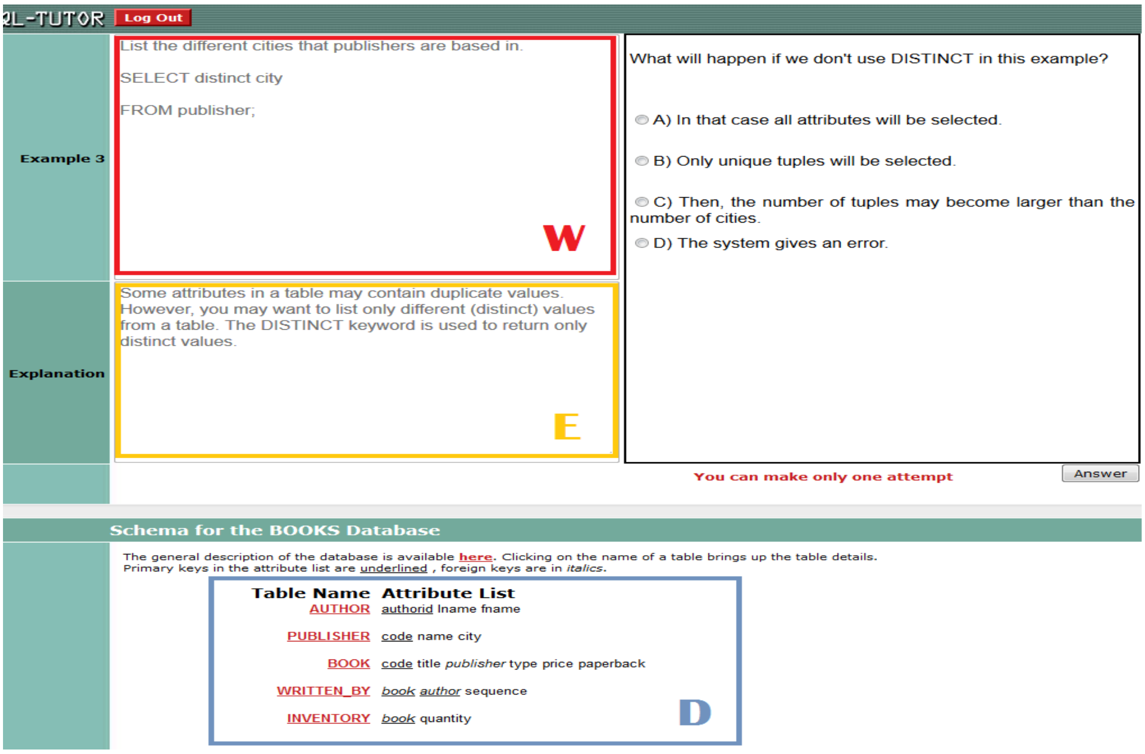

Figure 1. A screenshot of the WE mode of SQL-Tutor, with the three AOIs marked W, E and D

We extended the system by adding the worked-example mode (Shareghi Najar & Mitrovic, 2013; 2014). In this paper, we focus on how students study examples only. Figure 1 presents the screenshot of the WE mode, with a worked example at the top, followed by an explanation. We chose the *Books* database from the thirteen databases available in SQL-Tutor. The database schema is shown at the bottom of screen; primary keys are underlined, and foreign keys are in italics. Once a student confirms that s/he has finished studying the example (by clicking the button), the system presents a Procedural-focused Self-Explanation (P-SE) prompt (shown in the top right pane in Figure 1). The student is allowed only one attempt per P-SE prompt, which is a multi-choice question that encourages students to choose an answer that best explains solution steps. If the student selects a wrong answer, the system discloses the correct answer and lets the student continue with the following example. We added P-SE prompts after examples because previous research (Schwonke et al., 2009; Kim et al., 2007) shows that students acquire more conceptual knowledge than procedural knowledge from examples, so P-SE fosters students' procedural knowledge.

For this study, we made minor changes to the interface of the previous studies (Shareghi Najar & Mitrovic, 2013; Shareghi Najar, Mitrovic & McLaren, 2014). We added fixed gaps (> 30 pixels for the 1920*1200 resolution) between the prompt text and each of the options, in order to support identification of eye gazes. We also defined three Areas Of Interest (AOIs), which correspond to the parts of the interface which provide the worked example (W), explanation (E) and the database schema (D), as labelled in Figure 1. AOIs could be used to identify scanning sequences and transitions, and also to tally fixations (Goldberg & Helfman, 2010). Scrolling was not required and therefore the position of AOIs was fixed on the screen.

3. Study

The goal of our study was to identify meaningful differences in example processing between learners with different levels of SQL knowledge. The version of SQL-Tutor used in the study presented six WEs, each followed by a P-SE prompt. The participants were 22 students who also participated in our previous experiment (Shareghi Najar & Mitrovic, 2013), in which we used a different database. Thus, the examples used in this study were new to the participants. We collected the data using the Tobii TX300 eye tracker. Tobii allows unobtrusive eye tracking and collects data that can be analysed using Tobii Studio^{TM} or externally. After obtaining informed consent, we calibrated Tobii with students' eye gaze. Each student received NZ$20 voucher for participating in the study.

The length of each individual session was one hour. We have not administered pre/post-tests, as they would take 20-30 minutes and therefore leave very short time for learning. Instead, we used the pre/post test results from the previous study (Shareghi Najar, Mitrovic & McLaren, 2014), held only a week before the start of the current study. Therefore, our participants were not complete novices, as they have already learnt about SQL and used SQL-Tutor previously.

Our study is a quasi-experiment: the participants were not allocated to the two groups randomly, because we wanted to compare eye-gaze data for students with low and high existing SQL knowledge. It was not appropriate to determine the two groups solely on the basis of the pre/post test scores from the previous study, as the students have continued to use SQL-Tutor between the two studies.

In order to determine the two groups, we used the K-Medoids algorithm (Kaufman & Rousseeuw, 1987). The inputs for the clustering algorithm were the pre- and post-test scores from the previous study, and the P-SE scores and learning time from the current study. K-Medoids produced two clusters we labelled Weak Learners (WL) and Advanced Learners (AL), summarised in Table 1. The average scores of WLs on the pre-test, post-test and P-SE prompts are lower than the average for the whole group (the *Total* column), and they also spent less time studying examples. There are significant differences between the two groups on the pre-test, post-test and P-SE scores.

Table 1. Statistics about the two clusters (standard deviations provided in brackets)

	Total (22)	**WL** (12)	**AL** (10)	**p**
Pre-test (%)	40 (13)	33 (11)	48 (11)	<0.01**
Post-test (%)	70 (16)	63 (16)	79 (12)	0.02**
P-SE (%)	83 (13)	76 (11)	92 (9)	<0.01**
Time (min)	21.5 (9)	20 (8.6)	23 (9.8)	0.44

4. Results

Tobi Studio reports recording quality per session, which is the percentage of data samples which were valid (i.e. the samples where one or both eyes could be tracked). The overall recording quality for our study was 77%. We divided recordings into segments corresponding to individual examples, and calculated the segment quality. Then we excluded segments with the quality less than 40%, which resulted in elimination of the data collected for three participants. The overall recording quality for the remaining data was 90%. Additionally, one participant showed atypical behaviour, by studying examples only after receiving P-SE prompts, and therefore we eliminated this participant's data. The results reported are produced by analysing the data for the remaining 18 participants.

4.1 Analysis of Eye-Tracking Data

We extracted the following metrics from the Tobii Studio for each AOI:

- Individual fixation duration (seconds): duration of each individual fixation within an AOI;
- Total fixation duration (seconds): duration of all fixations within an AOI;

- Fixation count: the number of fixations on a particular AOI;
- Visit duration (seconds): the amount of time the participant spent looking at an AOI from when the eye gaze entered the AOI until the participant looked at a different AOI;
- Total visit duration (seconds): duration of all visits for a particular AOI;
- Visit count: the number of visits to an AOI.

We used the Mann-Whitney U test to compare the two groups on the reported metrics. Table 2 shows the results for D_{AOI}.

Table 2. Eye-gaze metrics for the database schema AOI

	WL (8)	**AL** (10)	**p**
Mean fixation duration (SD)	0.34 (.16)	0.52 (.24)	0.14
Mean total fixation duration (SD)	1.51 (1.52)	5.67 (5.23)	0.03**
Mean fixation count (SD)	6.75 (4.68)	27.50 (25.51)	0.02**
Mean visit duration (SD)	0.61 (.38)	2.29 (1.59)	0.01**
Mean total visit duration (SD)	1.60 (1.59)	6.79 (6.73)	0.03**
Mean visit count (SD)	3.75 (1.91)	6 (3.97)	0.20

There is no significant difference between the groups on fixation duration, but there are significant differences on the total fixation duration, fixation count, visit duration and total visit duration (p = .03, p = .02, p=.01, p = .03). That is, advanced learners fixated on D_{AOI} more than weak learners. Moreover, total visit duration shows that AL spent significantly longer time studying database schema than WLs.

Table 3. Eye-gaze metrics for the worked example AOI

	WL	**AL**	**p**
Mean fixation duration (SD)	1.44 (1.63)	1.42 (1.80)	0.63
Mean total fixation duration (SD)	56.10 (15.41)	70.00 (40.33)	0.89
Mean fixation count (SD)	230.75 (46.03)	282.40 (146.21)	0.76
Mean visit duration (SD)	21.98 (9.49)	17.89 (9.47)	0.36
Mean total visit duration (SD)	68.80 (21.47)	82.01 (45.88)	0.96
Mean visit count (SD)	24.38 (8.93)	32.60 (14.84)	0.15

Tables 3 and 4 present the results for W_{AOI} and E_{AOI}. There were no significant differences between the distributions of those metrics, showing that both WL and AL paid similar attention to those two AOIs.

Table 4. Eye-gaze metrics for the explanation AOI

	WL	**AL**	**p**
Mean fixation duration (SD)	1.19 (.17)	1.17 (.21)	0.90
Mean total fixation duration (SD)	28.30 (17.81)	26.82 (22.16)	0.63
Mean fixation count (SD)	130.25 (73.96)	117.50 (85.20)	0.79
Mean visit duration (SD)	10.82 (7.71)	8.43 (6.40)	0.46
Mean total visit duration (SD)	32.91 (20.50)	30.43 (24.69)	0.69
Mean visit count (SD)	19.75 (7.44)	20.40 (6.57)	0.76

4.2 Eye Gaze Pattern Analysis (EGPA)

We propose a new technique to analyse eye-gaze patterns named EGPA. Patterns are actions showing a student's attention on an AOI or eye gaze movements from one AOI to another, over a short time period (~1.5s). We identified four types of patterns: reading, mixed reading, transferring and scanning:

- *X* represents a reading pattern, and specifies that the student only looked at area X. For example, 'W' means that the student was reading worked example.
- *XyX* represents the mixed reading pattern, showing that the student had a short look at area Y while s/he was reviewing area X. For instance, 'EdE' shows that the student read the explanation (E), but s/he had a quick look at the database schema (D) while s/he was reading the explanation.
- *XY* represents a transferring pattern, when the student's eye gaze moved from area X to area Y. For example, 'WE' means that student's attention changed from the worked example to the explanation.
- S is the scanning pattern. This normally happens when a student sees the interface for the first time or when they are searching for information.

Table 5 reports the percentages of participants who used various patterns, and also the average pattern frequencies per group. The Mann-Whitney U test revealed no significant difference between the total number of patterns used by the two groups. The AL group used the D and ED patterns significantly and marginally significantly more often than WL (p = 0.03 and p = 0.08 respectively). The D pattern was used by 90% of advanced students compared to only 25% of WLs. The ED pattern was not used by weak learners at all, while half of advanced students have used it.

Table 5. Pattern statistics

	Students using patterns		**Average pattern frequency**		
	AL	WL	AL	WL	p
All patterns			18.60 (5.19)	18.75 (5.26)	0.97
W	100%	100%	4.8 (2.2)	5.25 (1.39)	0.83
E	90%	100%	2.2 (1.32)	2.375 (1.19)	0.83
D	90%	25%	1.1 (0.57)	0.375 (0.74)	0.03**
WeW	40%	63%	1.2 (1.81)	0.625 (0.52)	0.90
WdW	50%	25%	1.4 (1.84)	0.25 (0.46)	0.24
EwE	60%	75%	1.2 (1.32)	2.125 (2.1)	0.41
EdE	20%	38%	0.3 (0.67)	0.5 (0.76)	0.57
WE	90%	100%	3.5 (2.01)	4.625 (1.69)	0.24
WD	40%	25%	0.4 (0.52)	0.25 (0.46)	0.63
EW	50%	50%	0.7 (0.82)	0.875 (1.13)	0.90
ED	50%	0%	0.5 (0.53)	0	0.08*
DW	30%	0%	0.3 (0.48)	0	0.32
DE	0%	25%	0	0.25 (0.46)	0.41
S	70%	100%	1 (0.94)	1.25 (0.46)	0.41

4.3 Heat Maps and Gaze Plots

A heat map provides a two-dimensional graphical representation of eye-tracking data on a screen, in which colours are used to show fixation durations (Bojko, 2009). Humans can easily understand colours representing different temperatures, which makes heat maps easy to interpret: an area where most of eye gazes are fixated is visualised by using the red colour, and areas with lower levels of eye gazes range over yellow and green gradually. Heat maps also visualise eye gaze data over the screenshots that the participants see, which make them very easy to interpret.

Figure 2 shows the heat maps we generated from the eye-gaze data of all participants for examples 1 and 6. Overall, the students paid more attention to the AOIs in Example 1 than in the other examples. A possible interpretation is that students needed to familiarise themselves with the environment at the start of the session, and have therefore read all the information provided.

The heat map for Example 1 shows that when students looked at the database schema, they inspected most tables and their attributes. Example 6 shows the opposite eye-gaze behaviour, where the participants paid attention to only the relevant tables (BOOK and PUBLISHER).

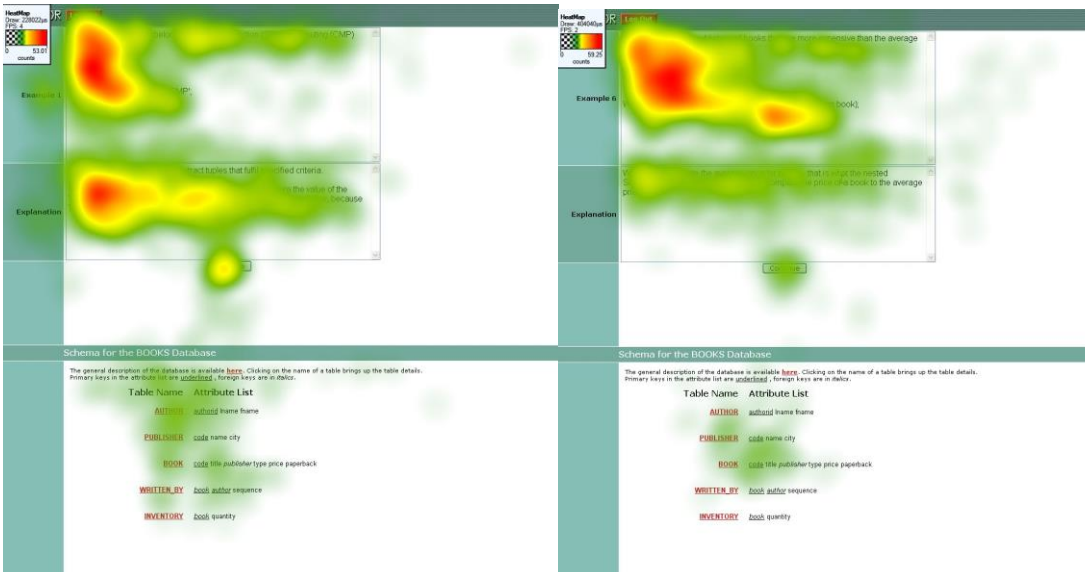

Figure 2. Heat maps for Example 1 (left) and Example 6 (right)

Figure 3 shows gaze plots of a typical advanced student on examples 1 and 6. The advanced student had a high number of fixations, and inspected all AOIs thoroughly. This participant read the problem statements and then studied the solutions. From the sequence of fixations, we can see that the advanced student looked at D_{AOI} when s/he was reading the problem statement. The information in the database schema is necessary in order to comprehend WEs. The advanced student examined all the tables during example 1; this helped the student to identify the primary keys, foreign keys and all information necessary for the example. The fixation numbers show that the advanced student read the explanation last.

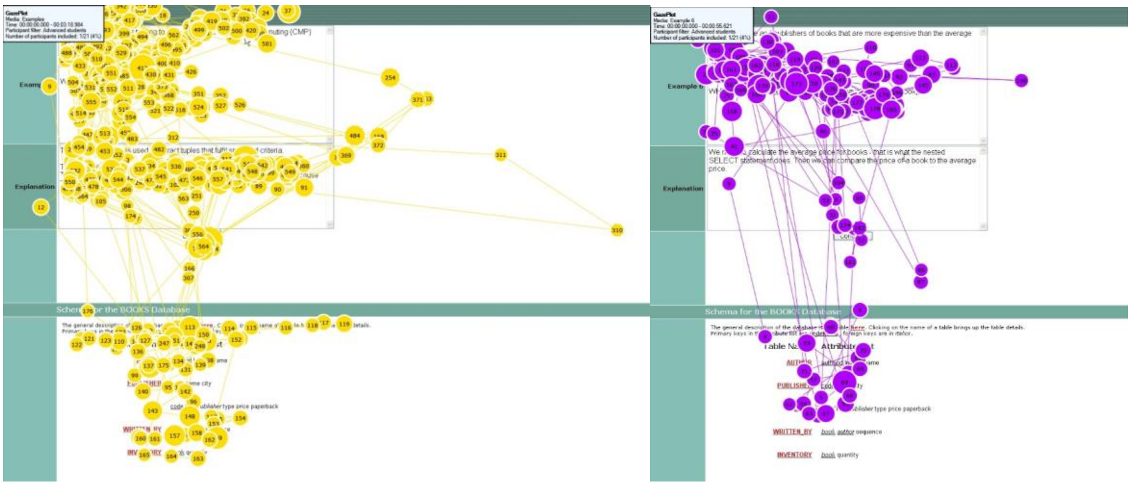

Figure 3. Gaze plots for an advanced student on examples 1 and 6

Figure 4 shows the gaze plots for the same examples but for a typical weak learner. It can be observed that the student pays much less attention to the AOIs in comparison to the advanced learner. The WL did not read all the information presented. For instance, Example 1 is about the BOOK table, but the weak learner did not fixate on any information about that table. The gaze plot for example 6 shows that the student has neither looked at the database schema nor at the explanation while reading that example.

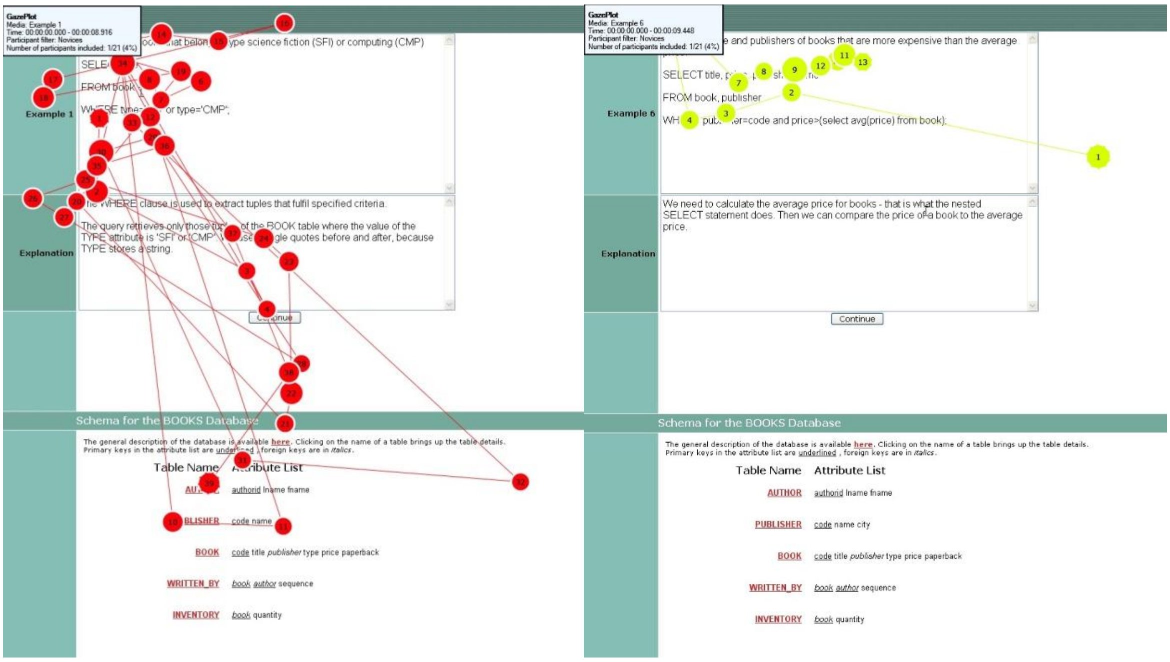

Figure 4. Gaze plots of a weak learner on examples 1 and 6

5. Conclusions

From our teaching experience, a good understanding of a database schema is critical in order to understand worked examples and solve problems. In the presented study, we used eye-tracking data to investigate whether there are differences in example processing between learners of varying existing knowledge. Overall, the results reveal that the advanced students paid more attention to the database schema than weak learners: advanced students used the D and ED patterns significantly and marginally significantly more often than their peers. Advanced students visited the database schema AOI more often and spent more time on it than the other group. The gaze plots for typical advanced and weak students show that they studied examples differently.

The presented results suggest that the ITS could provide hints to the weak learner in order to benefit more from WEs. The ITS could advise the student to examine the database schema during the first example, and to pay attention to primary/secondary keys especially. Later on during the session, the ITS could remind the learner to examine the information about the table(s) relevant to the example studied.

One of the limitations of our study is the small sample size. We plan to conduct larger studies, and also studies in areas other than SQL. Furthermore, the analyses performed were based on data captured over the whole session; therefore, the results may change when using data from a fragment of a session. It would also be interesting to observe how patterns change as students become more knowledgeable.

We plan to conduct a study to see whether or not drawing students' attention to the database schema will improve learning. We hypothesise that weak learners will learn more from examples with such adaptive guidance. Furthermore, it might be possible to use the student's eye-gaze behaviour in order to provide additional support. For instance, if the student's behaviour shows that the student has difficulties with the FROM clause (perhaps by having a very long fixation on the FROM clause), the system could then provide an explanation. Eye-gaze data may be further combined with the student model to provide adaptive examples.

References

Atkinson, R.K., Derry, S.J., Renkl, A., Wortham, D. (2000). Learning from examples: instructional principles from the worked examples research. *Review of Educational Research, 70*, 181–214.

Bull, S., Cooke, N., Mabbott, A. (2007). Visual attention in open learner model presentations: An eye-tracking investigation. In C. Conati, K. McCoy, & G. Paliouras (Eds.), *Proc. Int. Conf. on User Modeling* (pp. 187–196). Springer.

Bojko, A. (2009). Informative or misleading? heatmaps deconstructed. *Human-Computer Interaction. New Trends* (pp. 30–39).

Goldberg, J. H., Helfman, J. I. (2010). Comparing information graphics: a critical look at eye tracking. In *Proc. 3rd BELIV'10 Workshop: Beyond time and errors: novel evaluation methods for Information Visualization* (pp. 71–78).

Kardan, S., Conati, C. (2012). Exploring gaze data for determining user learning with an interactive simulation. In J. Masthoff, B. Mobasher, M. Desmarais, & R. Nkambou (Eds.), *Proc. Int. Conf. User Modeling, Adaptation, and Personalization* (pp. 126–138). Montreal, Canada.

Kaufman, L., Rousseeuw, P. (1987). Clustering by means of Medoids. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L_1–Norm and Related Methods* (pp. 405–416). North-Holland, Amsterdam.

Kim, R. S., Weitz, R., Heffernan, N. T., Krach, N. (2007). Tutored Problem Solving vs. "Pure" Worked Examples. In *Proc. 31st Annual Conf. of the Cognitive Science Society*. Austin, TX: Cognitive Science Society (pp. 3121–3126).

Kirschner, P.A., Sweller, J., Clark, R.E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Koedinger, K., Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.

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.), *Proc.11th Int. Conf. on Intelligent Tutoring Systems* (pp. 422–427). Chania, Greece: Springer Berlin Heidelberg.

McLaren, B., Isotani, S. (2011). When is it best to learn with all worked examples? In: G. Biswas, S. Bull, J. Kay, A. Mitrovic (Eds.) *Proc. 15th Int. Conf. Artificial Intelligence in Education* (pp. 222–229). Auckland, New Zealand: Springer Berlin Heidelberg.

Mitrovic, A. (1998). Experiences in implementing constraint-based modeling in SQL-Tutor. In B. Goettl, H. Halff, C. Redfield, & V. Shute (Eds.), *Proc. Int. Conf. on Intelligent Tutoring Systems* (pp. 414–423).

Mitrovic, A. (2003). An Intelligent SQL Tutor on the Web. *Artificial Intelligence in Education, 13*(2-4), 173–197.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25*, 258–266.

Shareghi Najar, A.,Mitrovic, A. (2013). Examples and tutored problems: how can self-explanation make a difference to learning? In H. C. Lane, K. Yacef, J. Mostow, P. Pavlik (Eds.), *Proc. 16th Int. Conf. on Artificial Intelligence in Education* (pp. 339–348).

Shareghi Najar, A., Mitrovic, A. Examples and tutored problems: is alternating examples and problems the best instructional strategy? *Research and Practice in Technology-Enhanced Learning, 9*(3), 439-459, 2014.

Shareghi Najar, A. S., Mitrovic, A., McLaren, B. M. (2014). Adaptive support versus alternating worked examples and tutored problems: which leads to better learning? In *Proc. 22nd Int. Conf. on User Modeling, Adaptation and Personalization* (pp. 171–182).

Sweller, J., Ayres, P., Kaluyga, S. (2011) *Cognitive load theory*. Springer.

Van Gog, T. (2011). Effects of identical example-problem and problem-example pairs on learning. C*omputers & Education, 57,* 1775-1779.