Chang, M. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

Towards Adaptive Provision of Examples During Problem Solving

Faiza TAHIR[a*], Antonija MITROVIC[a] & Valerie SOTARDI[b]
[a]*Intelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, New Zealand*
[b]*School of Educational Studies and Leadership, University of Canterbury, New Zealand*
*faiza.tahir@pg.canterbury.ac.nz

Abstract: Intelligent Tutoring Systems (ITSs) are effective in supporting learning, as shown in numerous studies. The goal of our project is to develop an adaptive strategy that would be capable of identifying situations during problem solving in which the student would benefit from worked examples. As a first step towards developing such a strategy, we conducted a pilot study in the context of SQL-Tutor, a mature ITS that teaches database querying. The participant could ask for a worked example whenever he/she wanted during problems solving. After each example, the participant specified whether the example was useful, and whether additional examples were needed. Participants' facial expressions and eye gaze were recorded. The findings show that the participants generally found examples useful, although in some cases they stated additional examples would be beneficial. The analysis of the eye gaze shows that students compared provided examples to their own solutions. Affect analysis shows that negative emotions reduced while engagement increased when participants viewed examples, and immediately after examples.

Keywords: intelligent tutoring system, problem solving, worked examples, eye tracking, affective modeling

1. Introduction

Intelligent Tutoring Systems (ITSs) have been proven to be very effective in supporting learning (Kulik & Fletcher, 2016; Mitrovic, 2012; VanLehn, 2011). The main activity in ITSs is problem solving, where the student receives help from the ITS adaptively, based on his/her actions and knowledge. On the other hand, there is a long tradition of research on learning from worked examples (WEs), starting from 1950s (Atkinson et al., 2000). A worked example contains the problem statement and a step-by-step solution with accompanying explanations. Atkinson et al. (2000) suggested the importance of worked examples in early stages of skill acquisition. Learning can also be increased when WEs are combined with self-explanation (Große & Renkl, 2007), problem solving (Cooper & Sweller, 1987), faded examples, (Renkl & Atkinson, 2003), or erroneous examples (Große & Renkl, 2007).

Examples have also been found beneficial when incorporated into ITSs. ELM-PE is one of the first ITSs to incorporate examples and their explanations (Burow & Weber, 1996). SE-Coach (Conati, Larkin, & VanLehn, 1997) guided students to self-explain examples; on the basis of student explanations and student model, it estimated the student understanding of a particular example. EA-Coach (Muldner & Conati, 2007) provided examples adaptively, based on learners' characteristics and example utility. Another study revealed the positive effects of providing WEs adaptively by fading their steps in a cognitive tutor (Salden et al., 2009). A study with SQL-Tutor compared learning from problems only, WEs only, or alternatively provided examples and problems to learners, found that a mixture of WE and problem solving resulted in best learning outcomes (Najar & Mitrovic, 2014). In follow-up studies, Najar and colleagues (2015) showed that adaptive selection of learning activities resulted in highest learning gains. Later on, erroneous examples were introduced in SQL-Tutor and proved to be helpful for advanced learners (Chen, Mitrovic & Mathews, 2019). Another study used a concept-based similarity approach to select most similar examples for the learner, when the learner fails to complete a Java program (Hosseini & Brusilovsky, 2017).

Most of these studies focused either on adaptive strategies for presenting WEs and/or problems, or on adaptive provision of example steps. However, there is a lack of research on adaptive strategies for providing examples to students when they need help during problems solving. In order to fill this gap, we designed and conducted a pilot study with SQL-Tutor (Mitrovic, 2003), the goal of which was to observe when and how students use worked examples during problem solving.

We start by presenting the worked example version of SQL-Tutor used in the pilot study, and then describe the procedure in Section 3. Section 4 presents the findings, while Section 5 presents conclusions.

2. Experimental Setup

The version of SQL-Tutor used in the pilot study contained ten problems. The screenshot in Figure 1 shows the problem-solving environment of SQL-Tutor. At the top of the page, there are several buttons allowing the student to change the database, select a problem, look at the history of the session or his/her student model, run the query, ask for help or exit the system. For each problem, there was one WE that was isomorphic to the problem, using the same database and same domain principles. Figure 1 shows a WE, which includes the problem statement, the solution accompanied with an explanation. After the explanation, the student was required to specify which clause of the Select statement he/she had difficulty with, and then to specify whether the example was useful, and whether additional examples were needed. The three questions were mandatory. In the study, we used the Tobii eye tracker to record the participant's eye gaze, and iMotions to record facial expressions.

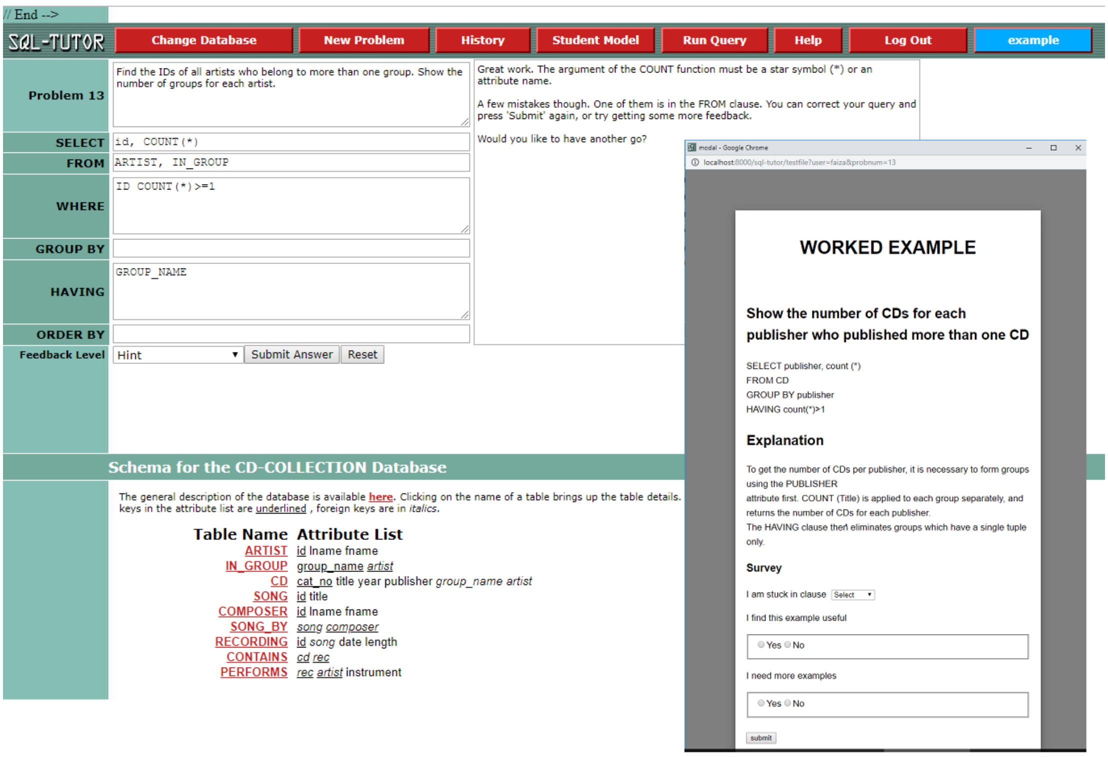

Figure.1 Screenshot of the worked example mode of SQL-Tutor

3. Procedure

We recruited seven undergraduate and three postgraduate students (two females, eight males), who were all studying Computer Science. Six participants were domestic students, while the remaining four were international (three Asian, and one Latin American). Five participants were aged 18-23, three 24-29 and two 30-35. All participants were familiar with SQL, and some of them have worked with

SQL-Tutor before the study. Each student had an individual session (40 minutes long), and was awarded a $20 voucher for their participation.

At the beginning of the study, the participants provided informed consent, and filled a short questionnaire, in order to collect basic information about participants and their level of familiarity with SQL-Tutor. The participant sat in front of the Tobii screen, and the standard Tobii calibration was completed. The calibration test took 6 seconds, and the experiment started only if results were excellent. Otherwise, the position of participant was readjusted and recalibration took place. The experimenter sat to the other side, and monitored the participant's face and eye gaze captured by both iMotions and Tobii. This monitoring ensured that during experiment full face of participant was captured so iMotions could record the facial features properly. The participants were instructed to solve at least five problems in SQL-Tutor, and to ask for examples as needed.

iMotions recorded participants' facial expressions, which needed to be post-processed first, and later converted into action units and emotions by using the Affectiva AFFDEX facial expression analysis engine. After post processing, only those recordings with the AFFDEX sampling rate quality higher than 80% were included in the analyses (i.e. in 80% of samples it was possible to identify facial features). Affectiva AFFDEX generated probabilistic estimates for the seven emotions (anger, disgust, surprise, sadness, joy, fear and contempt) based on macro-expressions (lasting 0.5-4 seconds) of each participant. We selected the amplitude-based thresholding technique to focus on the strongest emotion.

4. How Much have Participants used Examples?

Five participants have not used SQL-Tutor prior to the study. Two participants used SQL-Tutor a lot, while the remaining three had limited experience with the system. Table 1 shows how many participants attempted and completed each problem, asked for examples, and how much time was spent on average on the problem/example. The *Example* column specifies the number of participants who viewed examples. The participants mostly solved the problems in the provided order, from the easiest to the hardest. On average, participants attempted 6 problems (sd = 1.89). The four most difficult problems were attempted much less often, and no one completed problems 9 and 10. The participants completed 62% of the problems they attempted, and viewed examples in 59% of the cases. . For problems 1-5, as the problem complexity grows, the example use increases. In problems 2, 4, 5 and 10, participants viewed the examples more than once. When they viewed examples for the first time, they spent on average a minute viewing them. Upon the second and third viewing, this time decreased to 10-20 seconds only. The average time per example is proportional to the average time on problem.

Table 1

Problem, Example and Feedback Use; Time in minutes for problems, and in seconds for examples

Problem	Attempted by	Completed by	Time/ problem	Example	Time/ example	Feedback
1	9	9	1.48 (0.95)	2	40 (7.07)	1
2	8	7	4.18 (2.02)	7	41 (9.72)	3
3	9	7	1.6 (1.34)	1	25 (0)	2
4	10	6	5.58 (3.18)	7	72 (30)	4
5	9	4	6.5 (3.79)	6	46 (54)	5
6	5	3	3.46 (2.62)	3	38 (7.63)	2
7	2	1	2.1 (0.14)	2	16.5 (12)	0
8	3	1	2.05 (0.07)	2	26 (20)	1
9	3	0	2.2 (0.52)	3	30 (16)	0
10	3	0	7.3 (5.23)	3	43 (15.2)	0

The *Feedback* column of Table 1 shows the number of participants who have explicitly required specific levels of feedback (such as hint, partial/complete solution) while solving problems. More participants have used feedback for the easier problems (1-5) than for the rest of problems. This trend is opposite to how participants used examples.

Table 2

Participants' Opinions on Examples

Example	Viewed by	Useful	More examples
1	2	1	2
2	7	7	1
3	1	1	1
4	7	5	2
5	6	4	2
6	3	2	0
7	2	1	1
8	2	2	0
9	3	3	0
10	3	2	1

Table 2 shows participants' responses to the three questions given with WEs. In 78% of the cases participants found them useful, and in 36% of these cases, they wanted more WEs. For complex problems (problems 7-10), when completion rate was low (below 20%), the participants found examples very useful (80% of the cases), even when they have not completed those problems. This shows that regardless of success in problem solving, the participants found the examples useful. Please note that our study was voluntary, and therefore there was no need for students to complete all problems.

5. Eye Gaze Analysis

We analyzed the eye tracking data to determine how the participants read worked examples. Such analysis allows us to understand whether the participants use WEs appropriately. Each WE is isomorphic to the problem, and we expected students to compare the solution provided in the WE to their solution. The area of interest (AOI) was defined to cover the whole example (i.e. title, solution and explanation). The metrics included in eye tracking analysis are: (1) *Time in AOI*, i.e. the total time spent looking at the AOI; (2) *Visits*, i.e. the number of times the participant's eye gaze returns to the AOI; (3) *Fixation count*, showing the total number of fixations within the AOI; (4) Duration of the first fixation on the AOI; and (5) the average fixation duration in AOI. Table 3 shows the metrics for the ten examples (including multiple viewings) averaged over all participants who viewed those examples.

The *Time in AOI* column provides the average time spent by participants while examining a WE. The average number of visits to the AOI seems to increase as problems become more complex. As the number of example steps grows in later examples, the participants looked more often towards the problem solving area and schema.

Table 3

Averages (Standard Deviations) for Eye Tracking Metrics. Times are Reported in Seconds

Example	Time in AOI	Visits	Fixation count	First fixation duration (s)	Fixation duration
1	40 (2.8)	13 (1.4)	43 (6.3)	.2 (.07)	.19 (0)
2	33 (18.5)	9 (9.06)	38 (30)	.14 (.05)	.21 (.030)
3	24 (0)	5 (0)	7 (0)	.24 (0)	.23 (0)
4	59 (30)	27 (25)	132 (127)	.25 (.073)	.24 (.03)
5	59 (61.9)	18 (18.5)	111 (116.9)	.2 (.094)	.24 (.05)
6	51 (20.5)	23 (2.12)	82 (21)	.23 (.063)	.2 (.04)
7	23 (0)	13 (0)	51 (0)	.22 (0)	.25 (0)
8	49 (24.9)	32 (14)	107 (94)	.46 (.37)	.24 (.05)
9	34 (4.7)	10 (6.5)	55 (32)	.19 (.053)	.22 (.04)
10	47 (16.3)	20 (9.5)	107 (52)	.31 (.31)	.22 (.02)

The average fixation count shown in Table 3 is highest in examples 4, 5, 8 and 10. The highest fixation count shows that these participants did not just glance over those examples, but studied them thoroughly, not only the first time but also for the second or third viewing. The high fixation count and average duration of the first fixation on more complicated examples strengthens the above findings that as the number of example steps grows, more fixations were recorded.

6. Affect Analysis

Affectiva AFFDEX analyzes facial expressions and reports the values of seven emotions: anger, sadness, surprise, disgust, joy, contempt and fear, based on Ekman's (1999) categorization of emotions. However, these are general emotions, not the emotions specific to learning (Baker et al., 2010; Craig et al., 2004). Woolf and colleagues (2009) suggested mappings between Ekman's basic emotions to learning-related emotions: joy mapped to excitement, anger mapped to frustration, surprise mapped to boredom, and fear mapped to anxiety.

In line with the above mentioned research, we considered anger, joy, fear and surprise. We additionally included engagement, which is also crucial for learning (Craig et al., 2004; D'Mello, Picard, & Graesser, 2007). We observed some general trends. At the start of each problem, the dominant emotion was surprise, and once the problem was solved, the dominant emotion was joy. When participants received feedback from SQL-Tutor (upon submitting their solutions), the level of surprise was higher. In those situations when participants were able to solve the problem after receiving feedback, again the level of joy was increased. However, if they were not able to solve the problem, we noticed higher levels of anger, showing the participants' frustration.

Another event of interest is when students asked for examples. We analyzed the emotions for three different time intervals: (1) one minute before example use, (2) during example use, and (3) one minute after example use. Firstly, during one minute before participants asked for examples, the dominant emotions were anger and surprise, which seem to suggest that participants asked for examples when they were frustrated. Engagement increased and surprise decreased during or after working with examples. Fear was the least detected emotion; it decreased while and after working with examples and increased slightly when they were working on examples. Joy increased when they were working with examples, and immediately after, when the participants were able to complete problems after viewing examples. On the other hand, if the example did not help the participant solve the problem, we observed increased values for anger and surprise. In some of those cases, the participants asked for the example for the second time, and after that abandoned the problems. This is consistent with findings reported in the literature showing that frustration may lead to boredom, in which case learners loose interest in learning activities.

In summary, we found that participants asked for examples when the levels of anger (i.e. frustration) and surprise (i.e. anxiety) were elevated. Working with examples reduced such negative emotions and increased joy. After viewing examples, when participants turned again to problem solving, the intensity of negative emotions was low, but gradually increased if they were unable to solve the problem. The level of engagement increased for all participants during and after viewing examples. Therefore, examples have positive impact on participants' affective states, which will be helpful in learning with SQL-Tutor.

7. Conclusions

This paper presented the pilot study the goal of which was to analyze how participants use, study and feel about worked examples in their problem solving journey with SQL-Tutor. The results show that participants used examples extensively, particularly when the complexity of problems increased. Most participants agreed on the usefulness of examples and a few required more examples. This indicates the demand for examples during problem solving, regardless of success in problem solving. The eye gaze analysis revealed that participants tried to understand example structure by comparing examples with their solutions. Lastly, the positive impact of examples on participants' emotions is as examples reduced participant's negative emotions, and increased engagement and up to some extent joy.

The presented findings illustrate the need for and effectiveness of WEs, supported by cognitive and affective states of participants. These findings provide a starting point for developing an adaptive strategy for providing WEs adaptively, during problem solving. A limitation of our study is the small sample size. We plan to collect more data about how and when students use example in the forthcoming study in a large database course, which will enable us to develop and evaluate the adaptive strategy in follow-up studies.

Acknowledgements

We thank Jay Holland for helping with the study, as well as our participants.

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*(2), 181-214.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *Human-Computer Studies*, *68*(4), 223-241.

Burow, R., & Weber, G. (1996). *Example explanation in learning environments.* International Conference on Intelligent Tutoring Systems, pp. 457-465.

Chen, X., Mitrovic, A., & Matthews, M. (2019). Learning from Worked Examples, Erroneous Examples and Problem Solving: Towards Adaptive Selection of Learning Activities. *IEEE Transactions on Learning Technologies*. 10.1109/TLT.2019.2896080

Conati, C., Larkin, J., & VanLehn, K. (1997). A computer framework to support self-explanation. In *Proc. Int. Conf. Artificial Intelligence in Education* (Vol. 39, pp. 279-276).

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*(4), 347.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media*, *29*(3), 241-250.

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, *22*(4), 53-61.

Ekman, P. (1999). Facial expressions. *Handbook of cognition and emotion*, *16*(301), e320.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*(6), 612-634.

Hosseini, R., & Brusilovsky, P. (2017). A study of concept-based similarity approaches for recommending program examples. *New Review of Hypermedia and Multimedia*, *23*(3), 161-188.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, *86*(1), 42-78.

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, *13*(2-4), 173-197.

Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User modeling and user-adapted interaction*, *22*(1-2), 39-72.

Muldner, K., & Conati, C. (2007). Evaluating a Decision-Theoretic Approach to Tailored Example Selection. Proc. IJCAI, pp. 483-488.

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

Najar, A.S., Mitrovic, A., & Neshatian, K. (2015). Eye tracking and studying examples: how novices and advanced learners study SQL examples. *Journal of computing and information technology*, *23*(2), 171-190.

Najar, A. S., Mitrovic, A., McLaren, B. (2015) Examples and tutored problems: adaptive support using assistance scores. Proc. 24th Int. Joint conference on Artificial Intelligence, pp. 4317-4323.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational psychologist*, *38*(1), 15-22.

Salden, R. J., Aleven, V. A., Renkl, A., & Schwonke, R. (2009). Worked examples and tutored problem solving: redundant or synergistic forms of support? *Topics in Cognitive Science*, *1*(1), 203-213.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.

Woolf, B., Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *Int. Journal of Learning Technology*, *4*(3-4), 129-164.