Effect of Non-mandatory Use of an Intelligent Tutoring System on Students' Learning

Antonija Mitrović$^{(\boxtimes)}$ 🆔 and Jay Holland

Intelligent Computer Tutoring Group, University of Canterbury, Christchurch, New Zealand
`tanja.mitrovic@canterbury.ac.nz`

Abstract. Numerous controlled studies prove the effectiveness of Intelligent Tutoring Systems (ITSs). But what happens when ITSs are available to students for voluntary practice? EER-Tutor is a mature ITS which was previously found effective in controlled experiments. Students can use EER-Tutor for tutored problem solving, and there is also a special mode allowing students to develop solutions for the course assignment without receiving feedback. In this paper, we report the observations from two classes of university students using EER-Tutor. In 2018, the system was available for completely voluntary practice. We hypothesized that the students' pre-existing knowledge and the time spent in EER-Tutor, mediated by the number of attempted EER-Tutor problems, contribute to the students' scores on the assignment. All but one student used EER-Tutor to draw their assignment solutions, and 77% also used it for tutored problem solving. All our hypotheses were confirmed. Given the found benefits of tutored problem solving, we modified the assignment for the 2019 class so that the first part required students to solve three problems in EER-Tutor (without feedback), while the second part was similar to the 2018 assignment. Our hypothesized model fits the data well and shows the positive relationship between the three set problems on the overall system use, and the assignment scores. In 2019, 98% of the class engaged in tutored problem solving. The 2019 class also spent significantly more time in the ITS, solved significantly more problems and achieved higher scores on the assignment.

Keywords: Intelligent Tutoring System · Conceptual database design · Learning analytics · Voluntary practice · Learning effect

1 Introduction

Intelligent Tutoring Systems (ITSs) have been shown in controlled studies to produce significant improvements in learning in comparison to the classroom, e.g. [1–3]. Such randomized studies are usually based on the pre/post-test design, which allows for measuring learning gains. VanLehn [4] in his meta-review reported the effect size of $d = 0.76$ for ITSs, comparable to the effect sizes achieved in 1:1 human tutoring. Other recent meta-analyses of reported evaluations of ITSs show similar findings [5, 6].

But what happens when ITSs are available for voluntary practice? Existing literature suggests only a fraction of students typically engages with educational systems when their use is completely voluntary. For example, Gašević et al. [7] write that over 60% of

students are limited users of educational technology. Similarly, Denny and colleagues [8] report that only one third of students used PeerWise, a system that supports peer learning by allowing students to pose questions and to answer/rate questions written by their peers. Brusilovsky and colleagues [9] report that only one half of students engaged in voluntary practice in Python grids, a system that provides several types of activities for learning Python.

In this paper, we investigate the effect of EER-Tutor, a mature ITS that teaches conceptual database design. Different versions of EER-Tutor have been used in courses at the University of Canterbury since 2001. The system is available to students for voluntary practice, as a supplement to lectures and labs. The system has previously been evaluated in several studies, which proved its effectiveness. In this paper, we focus on two questions: how students use this ITS, as well as its effect on students' learning.

In Sect. 2, we briefly introduce EER-Tutor, while the following Section presents out hypothesized model. Section 4 presents the findings from the 2018 class. We then made a modification to the assignment, by requiring students to solve three problems in EER-Tutor in addition to a more open-ended problem, and developed a new hypothesized model. We present the findings from the 2019 class in Sect. 5. Finally, we reflect on the findings and discuss the limitations.

2 EER-Tutor

EER-Tutor is a mature ITS, that teaches conceptual database design using the Enhanced Entity-Relationship (EER) data model [10, 11]. Different versions of EER-Tutor have been available to students enrolled in a second-year relational database course since 2001. The system has also been used by numerous students world-wide[1].

We have presented the architecture, the student modeler and the adaptive features of EER-Tutor in previous papers [12–14]; here we briefly summarize its features necessary to understand the analyses we performed. Figure 1 shows a screenshot of EER-Tutor, with the text of the problem at the top, the drawing area in the middle pane, and the feedback area on the right. The student can select any problem he/she wants, or ask for a problem to be selected adaptively by the system (on the basis of the student model). The current version of the system contains 57 problems, which are ordered by their complexity. The student draws the diagram by selecting tools representing the components of the EER model, and names them by selecting words or phrases from the problem text. EER-Tutor highlights the names of created entity types in blue, the names of attributes in green and the names of relationships in magenta, thus providing an easy way for the student to see how much of the requirements have been covered. When the student submits the solution, EER-Tutor evaluates it and presents the feedback. In the situation shown in Fig. 1, the student specified the participation of the SENSOR entity type as partial (single line), while it should be total (double line). EER-Tutor highlights the relevant components of the solution in red to make it easier for the student to focus on the error.

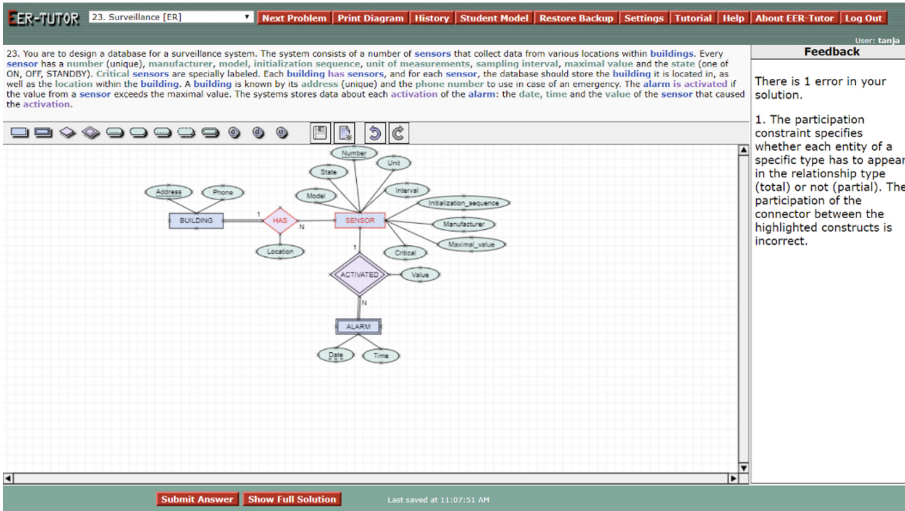[1] EER-Tutor was available on the Addison-Wesley's DatabasePlace portal from 2003 to 2016.

Fig. 1. The screenshot of EER-Tutor

We have implemented many versions of EER-Tutor, in order to evaluate some of its features, such as the open learner model [15–17], affect-aware animated pedagogical agent [18] and tutorial dialogues [19]. In all controlled studies, we have found significant improvements in learning.

3   2018 Database Course and Hypothesized Model

COSC265 is a single-semester (12 weeks) course on Relational database systems at the University of Canterbury, with three lectures and two lab hours per week. In 2018, there were 201 students enrolled in the course, who were completing Bachelor degrees in Computer Science (65%), Software Engineering (32%) or Information Systems (3%). Most of the students were in their second year, but there were 16 students repeating the course, and also some students taking the course in their first year (6%).

After a general introduction to databases (two lectures), the following four lectures were on conceptual database design using the EER model. At the end of the second week of the course (on July 27), the students were given an assignment worth 25% of the final grade, requiring them to develop an EER schema based on the given requirements. The assignment was due on August 24, which is the last day of week 6 (followed by a two-week break). Late submissions were allowed until August 31, in which case the students received a penalty of 15 marks.

EER-Tutor was introduced to the students briefly in a lecture in the second week, and the system was used in labs in the third week. The use of EER-Tutor was completely voluntary; the students did not receive any marks for solving problems in the ITS. The pre-test was given to students immediately after logging in, while the post test was given on a specific date. In addition to the 57 available problems, there is also a special mode of the tutor (referred to as mode 99), which allows students to draw EER diagrams without feedback. All students used

Fig. 2. Hypothesized path analytic model

this mode to draw their solutions for the assignment. The assignment was similar to the most complex problems in EER-Tutor. The final exam covered the whole course (50% of the final grade).
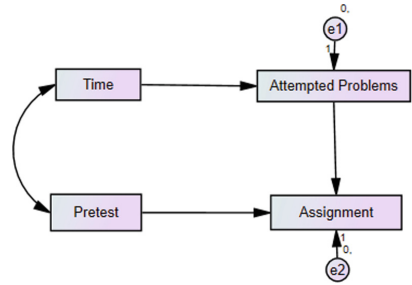
Figure 2 presents our path analytic model, based on previous research. Our first hypothesis is that the pre-existing knowledge (the pre-test score) will have a positive effect on the assignment score (*Assignment*). Positive correlation between pre-existing knowledge and the score after training (in our case the assignment score) is commonly found in the literature (e.g. [9]). Another common finding in the literature is that learning time is positively correlated with the final score. In our case, the time students spent in EER-Tutor was divided between working on assignment (i.e. drawing the diagram in mode 99) and tutored problem solving. The more time students spend in EER-Tutor, the more problems they attempt. We also hypothesize that attempted problems contribute to learning, as have been shown in previous studies with EER-Tutor. Therefore, the number of attempted problems mediates the relationship between time and the assignment score.

4   Findings from the 2018 Course

The pre-test contained seven questions (multiple choice or true/false), worth one mark each. Three questions asked to select correct definitions of EER concepts, while the remaining four questions required the student to select correct diagrams matching given requirements. There were two tests of similar complexities, which were randomly given to students as the pre-test. A student who received Test A as the pre-test, received Test B as the post-test and vice versa. Since there are two different tests used as the pre/post-test, we analysed the students' scores at the pre-test time, to make sure they were of similar difficulty. We report the statistics on pre-test scores in Table 1. There were 89 students who completed test A, and 86 completed test B. We found no significant difference between the pre-test scores on the two tests (t = 1.46, p = 0.15). The internal validity is acceptable for both tests, given the limited number of test questions and the broad range of tested knowledge [20].

Table 2 report statistics of how students interacted with EER-Tutor. The number of sessions and time are presented for 200 students, while the remaining rows present the values for the 153 students who attempted problem solving. One student never logged onto EER-Tutor. Forty-six students have only used the tutor to work on their

Table 1. 2018 pre-test scores

	Mean score (sd)	Cronbach's alpha
Test A (89)	3.62 (1.79)	0.59
Test B (86)	3.99 (1.56)	0.52

assignment. The median number of attempted problems is 13, while the median number of solved problems is 11. The median number of attempts per student was 34. For each submission, EER-Tutor provides feedback (as shown in Fig. 1). The pedagogical strategy implemented in the current version of EER-Tutor provides positive feedback on correct submissions, or provides up to three messages on mistakes when the submission is incorrect. The last row of Table 2 (*Hints*) reports the total number of error messages provided to the student.

Table 2. Summary statistics of EER-Tutor usage in 2018

	Min	Max	Median	Mean (sd)
Sessions	0	72	14	16.11 (11.84)
Time (min)	0	1285	125	189.85 (218.27)
Attempted problems	1	55	13	14.64 (10.47)
Solved problems	0	54	11	13.04 (10.09)
Attempts	1	364	34	49.95 (52.95)
Hints	0	614	43	75.58 (97.76)

Figure 3 (left) shows the number of students solving problems with EER-Tutor in weeks 2–7. Nineteen students started solving problems as soon as EER-Tutor was introduced in lectures in week 2. The highest number of students solving problems was recorded in week 3, when they used the ITS in the scheduled lab for the course. In weeks 5 and 6, 41 and 67 respectively worked on problems, in preparation for the assignment. There were few problems solved after week 7 until weeks 14–16, when students again solved problems in preparation for the exam. On average, students completed 78.55% of problems they attempted. Figure 3 (right) shows the number of students working on the assignment in weeks 3–7, with the average time (in minutes). Fourteen students started working on their assignment in the second week of the course. The peak in week 6 corresponds to the assignment deadline. Students who were late submitting the assignment used the system substantially in week 7.

Table 3 presents several performance measures. As EER-Tutor was available for voluntary practice, not all students started using it immediately, and consequently the date when students completed the pre-test ranged from July 23 to August 31. There were 16 students who either completed the pre- and post-test on the same day (because they started using the system late), or completed the two tests without attempting any
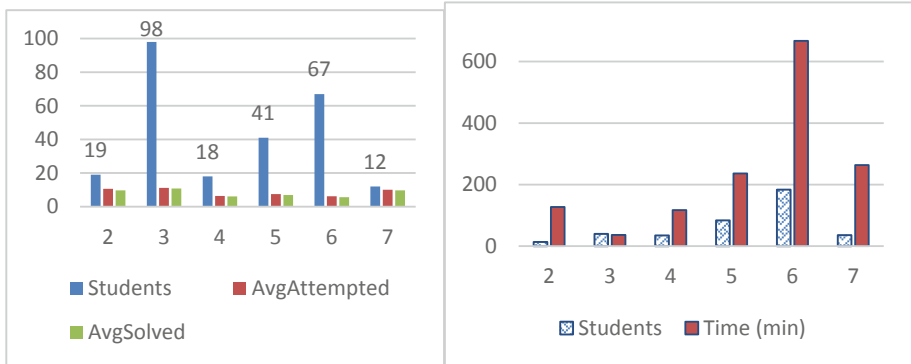
Fig. 3. Left: Number of students and attempted/solved problems. Right: Number of students and average time spent in mode 99

problems in between. For that reason, we did not include those students when calculating the normalized learning gain. Additionally, many students did not complete the post-test, so the number of students for whom we computed the normalized learning gain is 57. On average, the students achieved higher scores on the post-test compared to the pre-test scores, with the effect size (Cohen's d) of 0.38. One possible reason for the low value of the normalized gain is that students did not take the post-test seriously, as it did not contribute to the final grade. Additionally, the students were focused on completing their assignment at the time the post-test was administered.

Table 3. 2018 performance measures

	Min	Max	Median	Mean (sd)
Pretest % (180)	0	100	57.14	53.97 (24.33)
Normalized gain (57)	−1	1	0	0.17 (0.46)
Assignment % (198)	17	96	58.50	59.37 (13.20)

The path analytic model was evaluated with IBM SPSS Amos version 25, using the data collected from 179 participants for whom all relevant data were available (Fig. 4). The number of the parameters to be estimated in this model is 12. The amount of data we have is appropriate for this kind of analysis, as the recommendation is that there are at least ten participants per parameter [21]. All the variables in the path model are observed. Chi-square test (1.62) for this model
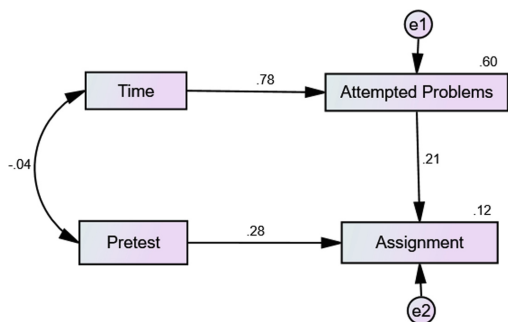

Fig. 4. 2018 model with path coefficients

(df = 2) shows that the model's predictions are not statistically significantly different from the data (p = .44). The Comparative Fit Index (CFI) was .99, and the Root Mean Square Error of Approximation (RMSEA) was .01. Therefore the model is acceptable: CFI is greater than .9 and RMSEA is less than .06 [22, 23]. All the path coefficients are significant at p < .005. Therefore, all our hypotheses are confirmed.

Therefore, tutored problem solving is important. One way to improve the performance of the class would be to require students to solve some problems in the ITS. In order to investigate how many problems make a difference, we divided the 2018 students post-hoc into two groups. The *Active* group contains those students who solved three or more problems in EER-Tutor. Table 4 reports the scores of Active students versus the rest of the class. There was a significant difference between the pre-test scores of the two subgroups of students (t = 2.32, p < .05). The Active students started with a higher level of knowledge, and used the system more, which may be the effect of those students being more motivated. There was no significant difference on the normalized gain, but the number of students who completed the post-test in both subgroups is small. This may show that the students have not taken post-test seriously; at that time of the course they were focused on completing their assignments, and taking a non-mandatory post-test was low priority. There were significant differences between the two subgroups on both assignment (t = 3.01, p < .005) and exam marks (t = 4.72, p < .001).

Table 4. Comparing the two 2018 subgroups

	Pretest	Norm. gain	Assignment	Exam
Active	57.02 (22.95) n = 120	0.18 (0.44) n = 51	61.44 (12.09) n = 127	73.97 (16.14) n = 121
Others	47.86 (26.01) n = 60	-0.03 (0.44) n = 19	55.68 (14.34) n = 71	62.37 (16.14) n = 67

5   2019 Course

Given the findings from 2018, we split the assignment into two parts for the 2019 class. The first part (*Assign1*) required students to solve three problems in EER-Tutor, without feedback. The chosen problems included one easy problem, and two problems of moderate difficulty. The hypothesized model is shown in Fig. 5. Similar to the 2018 model, we hypothesize that pre-existing knowledge and time spent in EER-Tutor will have a positive effect on the assignment score. The time students spent in EER-Tutor was divided between working on the three set problems in EER-Tutor (Assign 1), working on the second part of the assignment (i.e. drawing the solution using mode 99), and tutored problem solving. Therefore there are directional links between Time and Attempted problems, and Assign1. While working on Assign1, the students would improve their knowledge of database design; therefore we hypothesized a positive effect of Assign1 on Assign2. As in the previous model, we again hypothesize that the number of attempted

problems would have a positive effect on the second part of the assignment (*Assign2*). Assign1 mediates the relationship between the pre-test and attempted problems, as well as between pre-test and Assign2. The number of attempted problems mediates the relationship between the time spent in the system and Assign2, because students' knowledge would increase as they attempt problems in EER-Tutor.
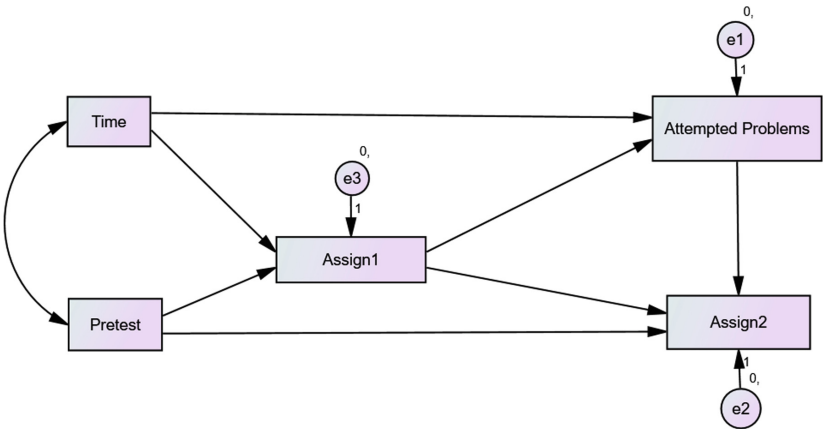

Fig. 5. Hypothesized model for the 2019 class

The only difference between the 2018 and 2019 instances of the course was in the assignment. The first part of the assignment was due at the end of week 4, while the second part was due at the end of week 6. There were 198 students enrolled in 2019, five of which have not engaged with the course at all. Out of the remaining 193 students, only one has not logged onto EER-Tutor. Table 5 presents some statistics of how students interacted with EER-Tutor. The number of sessions and the time in EER-Tutor are reported for 193 students, while the remaining rows of the table present the values for the 189 students who have attempted problem solving. Three students have not attempted problem solving, and used EER-Tutor solely to draw the solution for Assign2.

Table 5. Summary statistics of EER-Tutor usage in 2019

	Min	Max	Median	Mean (sd)
Sessions	0	74	15	17.22 (11.14)
Time (min)	0	2147	382	488.69 (355.89)
Attempted problems	1	57	22	22.11 (11.03)
Solved problems	0	50	18	17.53 (10.90)
Attempts	2	303	62	79.41 (60.93)
Hints	1	613	91	128.38 (117.91)

Table 6 presents the summary results about students' performance. Assign1 was worth 8% and Assign2 was worth 17% of the final grade. The last row in Table 6 presents the overall score for the assignment.

Table 6. 2019 performance measures (in percentages)

	Min	Max	Median	Mean (sd)
Pretest (184)	14.29	100	57.14	60.17 (21.03)
Assign 1 (188)	21	100	92	83.72 (16.29)
Assign 2 (189)	12	93	76.5	71.69 (15.32)
Assignment (192)	14	94	79.5	73.74 (16.39)

The estimated model is shown in Fig. 6. The model fits the data well, with CFI = 0.99, and RMSEA = 0. Chi-square test (1.71) for this model (df = 2) shows that the model's predictions are not statistically significantly different from the data (p = .43). All path coefficients are significant at p < .05, except *Pretest - >* Assign1 (p = .077).
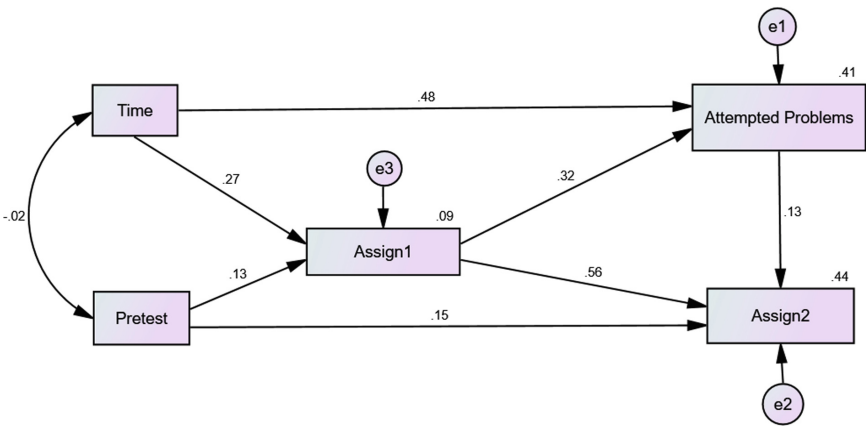

Fig. 6. 2019 model with standardized path coefficients

For the reader's convenience, we present the 2019 data on weekly use of EER-Tutor together with the 2018 data in Fig. 7. In 2019, students used EER-Tutor for the first time in week 3, and therefore we present the data for weeks 3 to 7 only. Many more students engaged in tutored problem solving in weeks 3 and 4 in 2019 in comparison to 2018. We believe the reason for that is the requirement for three problems to be solved in 2019 by week 4, which motivated students to practice more. In 2018, students have spent more time in mode 99 (working on the assignment) than in 2019; that might be because the 2019 students learnt more from tutored problem solving in early weeks and were therefore able to complete the assignment faster.
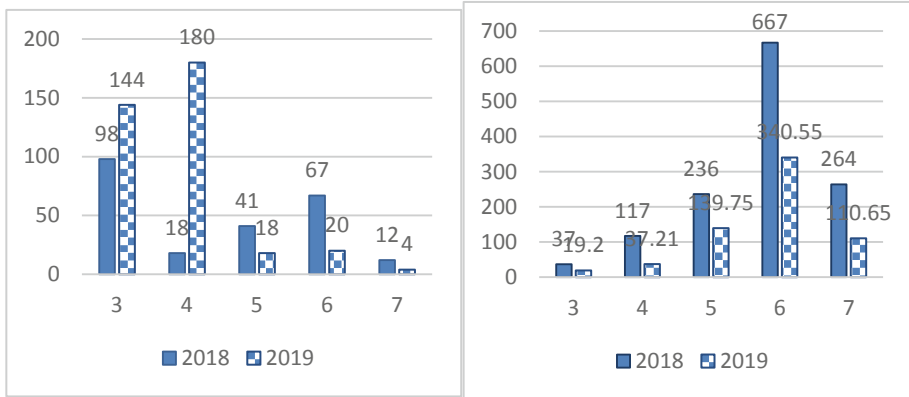
Fig. 7. Left: the number of students solving problems in EER-Tutor; Right: time spent in mode 99

6 Conclusions

In this paper we reported how students used EER-Tutor for voluntary practice in two consecutive years of the same course. Our findings are in contrast to the finding from the literature which shows that many students (50% or more) do not engage in voluntary practice with educational technology [7–9]. On the contrary, in our 2018 cohort, 23% of students used the tutor only to draw their assignments, and have not attempted any problem solving. The majority of the class (77%) used EER-Tutor both to work on the assignment, and for tutored problem solving.

One of the reasons for limited use of educational technology reported in the literature is the low levels of self-regulation skills and motivation [9, 24]. Since tutored problem solving in EER-Tutor was voluntary, it may be the case that the students who solved a lot of problems are more motivated students. We did, however, find that the number of attempted problems and time spent with EER-Tutor are significant predictors of their performance on the assignment. Students who solved at least three problems in EER-Tutor in 2018 received significantly higher marks on the assignment than the rest of the class.

Therefore, one straightforward recommendation for improving students' learning is to introduce a degree of mandatory problem solving. We have made that change in 2019, when the students were required to solve three problems in EER-Tutor as the first part of the assignment. In 2019, only three students have used EER-Tutor solely to draw the EER diagram; therefore, the percentage of students who used EER-Tutor for tutored problem solving increased from 77% in 2018 to 98% in 2019. In 2018, 69.5% of students solved at least one problem in EER-Tutor, while in 2019 that percentage increased to 91.71% (this is one problem in addition to the three mandatory problems). Therefore, requiring students to solve three problems increased their voluntary use of EER-Tutor in 2019. Comparing the two classes, we found that the 2019 class spent significantly more time in the tutor ($t = 10.03$, $p < .001$), solved significantly more problems in EER-Tutor ($t = 7.03$, $p < .001$) and achieved significantly higher marks on the assignment ($t = 9.52$, $p < .001$). Comparing the 2018/2019 assignment scores may not be fair, as the

two assignments may not have been of the same complexity, but the other two measures (time and the number of solved problems) show evidence that the intervention (requiring students to solve three prescribed problems) has made a difference.

One limitation of our study is that we have not collected data about students' self-regulation skills and motivation. We plan to collect such data in the 2020 class, which will allow us to look deeper into individual differences.

References

1. Anderson, J.R., Corbett, A., Koedinger, K., Pelletier, R.: Cognitive tutors: lessons learned. J. Learn. Sci. **4**(2), 167–207 (1995)
2. VanLehn, K.: The behavior of tutoring systems. Artif. Intell. Educ. **16**(3), 227–265 (2006)
3. Mitrovic, A.: Fifteen years of constraint-based tutors: what we have achieved and where we are going. User Model. User-Adap. Inter. **22**(1–2), 39–72 (2012)
4. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ. Psychol. **46**(4), 197–221 (2011)
5. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: a meta-analysis. J. Educ. Psychol. **106**(4), 901 (2014)
6. Kulik, J.A., Fletcher, J.: Effectiveness of intelligent tutoring systems: a meta-analytic review. Rev. Educ. Res. **86**(1), 42–78 (2016)
7. Gašević, D., Mirriahi, N., Dawson, S.: Analytics of the effects of video use and instruction to support reflective learning. In: Proceedings of the 4th International Conference on Learning Analytics and Knowledge, pp. 123–132. ACM (2014)
8. Denny, P., McDonald, F., Empson, R., Kelly, P., Petersen, A.: Empirical support for a causal relationship between gamification and learning outcomes. In: Proceedings of CHI Conference on Human Factors in Computing Systems, p. 311. ACM (2018)
9. Brusilovsky, P., Malmi, L., Hosseini, R., Guerra, J., Sirkiä, T., Pollari-Malmi, K.: An integrated practice system for learning programming in Python: design and evaluation. Res. Pract. Technol. Enhanced Learn. **13**(1), 18 (2018)
10. Chen, P.: The entity relationship model - toward a unified view of data. ACM Trans. Datab. Syst. **1**, 9–36 (1976)
11. Elmasri, R., Navathe, S.B.: Fundamentals of Database Systems. Addison Wesley (1994)
12. Suraweera, P., Mitrovic, A.: KERMIT: a constraint-based tutor for database modeling. In: Cerri, S.A., Gouardères, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 377–387. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-47987-2_41
13. Suraweera, P., Mitrovic, A.: An intelligent tutoring system for entity relationship modelling. Artif. Intell. Educ. **14**(3–4), 375–417 (2004)
14. Zakharov, K., Mitrovic, A., Ohlsson, S: Feedback micro-engineering in EER-Tutor. In: Looi, C.-K., McCalla, G., Bredeweg, B., Breuker, J. (eds.) Proceedings of International Conference Artificial Intelligence in Education. IOS Press, pp. 718–725 (2005)
15. Hartley, D., Mitrovic, A.: Supporting learning by opening the student model. In: Cerri, S.A., Gouardères, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 453–462. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-47987-2_48

16. Duan, D., Mitrovic, A., Churcher, N.: Evaluating the effectiveness of multiple open student models in EER-Tutor. In: Wong, S.L. et al. (eds.) Proceedings of 18th International Conference on Computers in Education, pp. 86–88 (2010)

17. Mathews, M., Mitrovic, A., Lin, B., Holland, J., Churcher, N.: Do your eyes give it away? Using eye tracking data to understand students' attitudes towards open student model representations. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 422–427. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30950-2_54

18. Zakharov, K., Mitrovic, A., Johnston, L.: Pedagogical agents trying on a caring mentor role. In: Luckin, R., Koedinger, K., Greer, J. (eds.) Proceedings of 13th International Conference on Artificial Intelligence in Education, Los Angeles, pp. 59–66 (2007)

19. Weerasinghe, A., Mitrovic, A., Martin, B.: Towards individualized dialogue support for ill-defined domains. Artif. Intell. Educ. **19**(4), 357–379 (2009)

20. Taber, K.S.: The use of Cronbach's alpha when developing and reporting research instruments in science education. Res. Sci. Educ. **48**(6), 273–1296 (2018)

21. Raykov, T., Widaman, K.F.: Issues in applied structural equation modeling research. Struct. Eqn. Model. **2**(4), 289–310 (1995)

22. Hu, L.T., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Eqn. Model.: Multidiscip. J. **6**(1), 1–55 (1999)

23. Teo, T., Tsai, L.T., Yang, C.C.: Applying structural equation modeling (SEM) in educational research: an introduction. In: Application of Structural Equation Modeling in Educational Research and Practice, pp. 1–21. Brill Sense (2013)

24. Lust, G., Elen, J., Clarebout, G.: Regulation of tool use within a blended course: Student differences and performance effects. Comput. Educ. **60**(1), 385–395 (2013)