Metacognitive Practice Makes Perfect: Improving Students' Self-Assessment Skills With an Intelligent Tutoring System

Ido Roll[1], Vincent Aleven[2], Bruce M. McLaren[2] and Kenneth R. Koedinger[2]

[1] University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[2] Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
ido@phas.ubc.ca, {aleven, bmclaren, koedinger}@cs.cmu.edu

Abstract. Helping students' improve their metacognitive and self-regulation skills holds the potential to improve students' ability to learn independently. Yet, to date, there are relatively few success stories of helping students enhance their metacognitive skills using interactive learning environments. In this paper we describe the Self-Assessment Tutor, an intelligent tutoring system for improving the accuracy of the judgments students make regarding their own knowledge. A classroom evaluation of the Self-Assessment Tutor with 84 students found that students improved their ability to identify their strengths while working with the Self-Assessment Tutor. In addition, students transferred the improved self-assessment skills to corresponding sections in the Geometry Cognitive Tutor. However, students often failed to identify their knowledge deficits a-priori and failed to update their assessments following unsuccessful solution attempts. This study contributes to theories of Self-Assessment and provides support for the viability of improving metacognitive skills using intelligent tutoring systems.

Keywords: Metacognition, self-regulated learning, intelligent tutoring systems, Self-assessment, help seeking, feeling of knowing (FOK).

1 Introduction

Students who apply productive metacognitive and self-regulation skills show better learning when working with interactive learning environments [1]. Therefore, many tutoring systems support various self-regulation skills [2,3,4]. Yet, only few systems attempt to improve students' self-regulation skills in a manner that persists even after support is removed and transfers to new learning situations. Two success stories are Betty's Brain, a learning-by-teaching environment for scientific concepts [5], and the Help Tutor, an add-on tutoring agent that gives metacognitive feedback on students' help-seeking behaviors while learning Geometry [6]. In both cases, students who received metacognitive prompts [5] or feedback [6] improved corresponding aspects of their learning trajectories in unsupported transfer tasks within the same environments.

In the current paper we describe the Self-Assessment (SA) Tutor, an intelligent tutoring system for improving students' SA skills. The term SA refers to students' tendency and ability to accurately evaluate their knowledge while learning [7, 8]. Accurate SA was shown to correlate with productive help-seeking behaviors [9]. A small number of systems provide support for SA in order to help students choose appropriate cognitive strategies [10] and monitor their progress [11]. In order for students' SAs to be accurate, students should be aware of the relative strengths and weaknesses of their knowledge, in relation to a target task [12]. However, students often over-estimate their ability [7]. Students who lack sufficient domain knowledge are especially likely to make inaccurate SAs, probably because they cannot distinguish between correct and incorrect answers, even when the solutions are presented to them [12]. In fact, students often base their assessments on familiarity with the problems, not with the answers [13].

The current study further evaluates the relationship between domain knowledge and accuracy of SAs, and focuses on acquisition, calibration, and transfer of SA skills in the context of an interactive learning environment. Specifically, we address the following questions:

1   Do students who lack domain knowledge also make less accurate SAs?
2   How do students use their actual problem-solving ability to calibrate their SAs?
3   Does the SA Tutor help students improve the accuracy of their SAs?
4   Do improved SA skills transfer to unsupported sections of the problem-solving environment?

In what follows we describe the SA Tutor and its classroom evaluation.

2   The Self-Assessment Tutor

The goals of the SA Tutor are to help students get in the habit of assessing their ability, improve the accuracy of their SAs, and use their SAs to inform strategy choice. The SA Tutor, an intelligent tutoring system [14], adheres to several principles of metacognitive tutoring [15]. The SA Tutor is a *learning by doing* environment in that students learn to self assess by practicing SA in the context of math problem solving. The SA Tutor helps students set the following subgoals: predict one's own ability, attempt to solve the problem, reflect on the experience, and plan future interaction. [15]. Since students who identify their own errors learn better than students who receive feedback on their errors [16], the SA tutor helps students to identify their SA errors. Adaptive feedback is given to students who fail to attend mismatches between their SAs and their actual performance. Last, the SA Tutor supports the entire problem-solving process, starting before students attempt to solve the target problem, and ending after students reflect on the solution.

Students begin the SA process by predicting whether they could solve a given target problem without making errors (Question 1 in Figure 1). Students reply by choosing either "yes" or "no, I need a hint," in which case a relevant hint is displayed. Both replies are legitimate, and no feedback is given on students' initial SA. Students are then asked to solve the target problem (Question 2). On this step, typical support is available (correctness feedback, error messages, and on-demands hints). Question 3
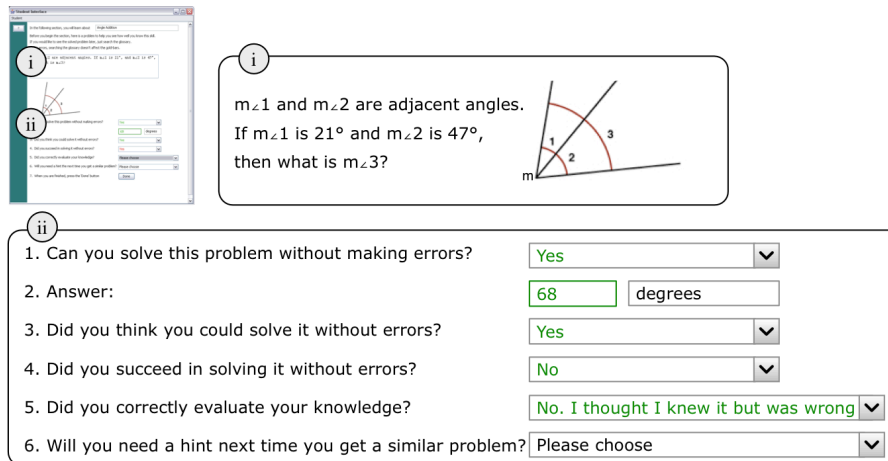
Fig 1. The SA Tutor (top left corner) includes two components: (i) domain-level problems, and (ii) self-assessment scaffold.

asks students to recall their initial SA and Question 4 asks students to reflect on whether they solved the target problem without making errors. Feedback on questions 3 and 4 is given to insure accurate recollection of students' initial SA and actual ability. Question 5 is key in getting students to compare their initial SA to their actual ability. In response to the question "did you correctly evaluate your knowledge?", students can choose "yes", "no--I thought I knew it but was wrong", or "no--I knew more than I predicted". Feedback on this question is contingent on students' initial SA and actual ability. For example, a student who estimated she could solve the target problem, yet failed to do so without errors, is expected to choose "no—I thought I knew it but was wrong". Last, students predict the need for help on new, similar, problems, by choosing either "yes, I will need the advice", or "no, I think I got it" (Question 6). No feedback is given on this question. The SA Tutor is an example-tracing tutor and was built using the Cognitive Tutor Authoring Tools [17].

In our study, the SA Tutor was used in conjunction with the Geometry Cognitive Tutor. Each section of the SA Tutor includes 3-5 problems, each of which targets a specific skill that is practiced in the subsequent section of the Geometry Cognitive Tutor. Students first evaluate their ability on the target set of problems in the SA Tutor. Students then complete a sequence of problems that require the same skills, using the Geometry Cognitive Tutor.

3   Methods

The SA Tutor was evaluated in a classroom study together with the Help Tutor [6]. An analysis of students' help-seeking behaviors is presented elsewhere [6,18].

Participants: The study took place in a rural vocational high school with 84 students in five classrooms, taught by two teachers. All students, 10th and 11th graders, were enrolled in the Cognitive Tutor Geometry class, and thus were familiar

with the Cognitive Tutor and its interface. Because the experimental conditions differed substantially, whole classes were assigned to conditions, balancing, across conditions, the number and level of students. 46 students in three classes were assigned to the SA Condition, while 38 students in the remaining two classes were assigned to the Control Condition

Materials: Students in both conditions worked on two units from the Geometry Cognitive Tutor: Angles (Unit 1) and Quadrilaterals (Unit 2). Each of the units had a single warm-up problem, followed by 3 sections. Each section focused on a different set of skills within the general topic of the unit. Students in the Control condition worked with the unmodified Geometry Cognitive Tutor, which did not include the SA Tutor or the Help Tutor. Students in the SA Condition alternated between the SA Tutor and the Geometry Cognitive Tutor augmented with the Help Tutor.

Procedure: The study spanned 3 months. During Month 1 all students worked on Unit 1 in their respective conditions. During Month 2 the study was put on hold while students prepared for statewide exams using the unmodified Geometry Cognitive Tutor. During Month 3 students worked on Unit 2, again according to the conditions to which they had been assigned. All students were assigned to the beginning of each unit at the same start date. Progress within each unit was at an individual pace. Figure 2 illustrates the structure of the study.

Month:	Month 1 – Unit 1, Angles.							Month 2 – Various CT units	Month 3 – Unit 2, Quadrilaterals.						
Section:	Warm -up	Section 1.1		Section 1.2		Section 1.3			Warm -up	Section 2.1		Section 2.2		Section 2.3	
SA:	CT	SA	CT	SA	CT	SA	CT		CT	SA	CT	SA	CT	SA	CT
Control:	Cognitive Tutor								Cognitive Tutor						

Fig 2. The procedure of the study. SA and CT denote SA Tutor and Cognitive Tutor respectively.

Analysis: Unless stated otherwise, all analysis involves students in the SA Condition only. Question numbers refer to the questions in the SA Tutor as shown in Figure 1.

4   Results

On average, students worked with the SA Tutor for 18 minutes. As it turns out, many students took longer than expected to complete sections 1.2 and 2.1 in the Geometry Cognitive Tutor, and thus did not reach the more advanced sections. In Unit 1, all 46 students worked on Section 1.1, 37 students worked on Section 1.2, and only 14 students reached Section 1.3. In Unit 2, 44 students worked on Section 2.1, and only 12 and 2 students reached Sections 2.2 and 2.3 respectively.

Research Question 1: Effect of Domain Knowledge. The SA Tutor asks students to predict their ability to solve a target item (Question 1), and following their prediction, to solve it (Question 2). Overall, students assessed their ability correctly on 77% of all problems. The accuracy of students' assessments depends on their knowledge level.

There is a high correlation between having the relevant domain knowledge (as assessed by averaging performance on Question 2 on all items within each section) and making accurate SA on the same set of items; $r(160) = .52$, $p < .0005$. The relationship between having relevant domain-level knowledge and accuracy of SA is most apparent when looking at the single item level (Table 1). Students who had sufficient knowledge to solve the target item predicted their success (prior to attempting) on 84% of the items, while students who lacked sufficient knowledge to solve the target item predicted their failure (prior to attempting) only on 37% of the items. Thus, over-estimation was much more common than under-estimation.

Table 1. Initial SA vs. competence (number of items and row-based percentage).

		Students' initial SA (Question 1)		
		Already know	Need help	Overall
Students' ability to solve the target item (Question 2)	High	455 (84%) ✓ True positive	85 (16%) ✗ False negative (under-estimation)	540 (100%)
	Low	64 (63%) ✗ False positive (over-estimation)	37 (37%) ✓ True negative	101 (100%)

Research Question 2: Calibration of SA. The SA Tutor asks students to report their SA twice for each skill: once before solving the target item (Question 1) and once after solving it (Question 6). Therefore, students can use their performance on the target item (Question 2) to calibrate their SA.

A repeated measures ANOVA (with initial- and updated-SA as a time series, and actual performance as a treatment) found that updated SA (Question 6) depends on the interaction between initial-SA (question 1) and actual performance (Question 2), $F(1,638) = 36$, $p < .0005$. As Table 2 shows, students' updated SA (Question 6) relies heavily on their initial SA (Question 1), but was fine-tuned based on their actual performance (Question 2). The significant interaction shows that students who under-estimated their ability updated their SAs more often than students who over-estimated their ability. In fact, 77% of the students who thought they already knew how to solve the item did not update their SAs following their failure to solve the item. The high persistence of over-estimation is especially noteworthy, given that a single failure is sufficient to suggest that the student does not possess sufficient knowledge.

Table 2. Updated SA: Students' reported confidence in their ability to solve additional problems that require the same skills without additional assistance (Question 6).

Initial SA (Q1)	Actual Performance (Q2)	Updated SA (Q6)
Already know	Got it right	88% will not need additional help
	Got it wrong	77% will not need additional help
Need Help	Got it right	51% will not need additional help
	Got it wrong	27% will not need additional help

Research Question 3: Metacognitive Improvement. Due to the high attrition, and to avoid a selection bias (in that data in the advanced sections pertains to better students), we evaluate the improvement in students' SA only on sections in which attrition was low: Unit 1 Sections 1.1 and 1.2, and Unit 2 Section 2.1. Overall, students became more accurate in their initial self-assessments, as evaluated by comparing their SAs on Question 1 to their actual performance on Question 2: Section 1.1: 71%; Section 1.2: 74%; section 2.1: 79%. However, a likely explanation is that students' SAs improved since their domain-knowledge increased. To control for the effect of domain learning, we analyzed the accuracy of students' SAs separately for items on which students had sufficient knowledge and items for which students lacked sufficient knowledge (as evaluated by performance on Question 2 in each problem). An ANOVA of accuracy vs. section, using data from items that students solved without errors (high competence items), found that students improved their SA significantly from Section 1.1 (77%) to Section 1.2 (88%): $F(1,294) = 4.8$, $p < .03$. There was also a positive trend from Section 1.1 to 2.1 (83%) on high-competence items ($p = .13$). However, there was no improvement in the accuracy of students' SAs on items that they subsequently failed to solve correctly (Section 1.1: 37%; Section 1.2: 40%; Section 2.1: 39%). These results suggest that students got significantly better at identifying their strengths, but not their weaknesses.

Research Question 4: Transfer of SA skills. To evaluate whether students transferred their improved ability to self-assess to an unsupported learning environment, we compare students' SAs in the SA Tutor to their actual help-seeking behavior in the Geometry Cognitive Tutor. Specifically, we examine the rate of asking for help in the Cognitive Tutor prior to attempting new problem-steps. One expects to see that students seek more help in the Geometry Cognitive Tutor on skills for which they report to have low SA in the SA Tutor. It is only natural that students ask for more help on skills they do not know. However, as shown earlier, students are relatively poor at identifying their limitations.

The correlation between skills on which students sought more help in the Cognitive Tutor and skills on which students reported to have low initial-SA in the SA Tutor is high and significant, $r(7) = .75$, $p = .02$. Other factors such as inherent difficulty or generic SA skills may affect students' help-seeking behaviors within the Cognitive Tutor. These factors can be accounted for by partialling-out the corresponding help frequencies on the same skills of students in the Control Condition, who were susceptible to the same factors, yet did not work with the SA Tutor. The partial-correlation between help-requests in the Cognitive Tutor and reported need for help in the SA Tutor, controlling for help-requests in the Cognitive Tutor by students in the Control Condition, remains high and significant: *partial-r*(6) = .73, $p = .04$. This suggests that training within the SA Tutor, rather than item difficulty or generic self-assessment skills, accounts for the high correlation between students' SA and help-seeking behavior.

Discussion and Summary

We have described the SA Tutor, an intelligent tutoring system for SA. The SA Tutor scaffolds the SA process in four steps: predicting ability to solve a target problem; attempting to solve that problem; reflecting on the SA by comparing the initial SA to the actual performance; and updating the SA for future interaction.

A classroom evaluation of the SA Tutor found that the SA Tutor helped students improve several aspects of their SA behavior. The SA Tutor helped students improve the accuracy of their initial SA with practice, and students also calibrated their SAs based on their actual performance on the target set of problems. Last, analysis of students' help-seeking behaviors in the Geometry Cognitive Tutor indicates that students transferred their improved SA knowledge to the subsequent sections in the Geometry Cognitive Tutor.

The *in-vivo* study emphasizes the large dependency of SA on domain knowledge, as was previously found in the lab [12]. Students' SAs were accurate on 84% of the items that they subsequently solved correctly, compared with a mere 37% of the items that they subsequently failed to solve. The rate of over-confidence did not decline with practice, while the rate of under-confidence declined significantly. The phenomenon of over-confidence seems not only common, but also persistent. Students were the least likely to use evidence to calibrate their assessment on items on which they were over-confident, even in the presence of feedback (see Table 2). Apparently, students did not attribute their failure to solve the problem to lack of relevant knowledge. As stated by Kruger and Dunning, "those with limited knowledge in a domain suffer a dual burden: Not only do they reach mistaken conclusions and make regrettable errors, but their incompetence robs them of the ability to realize it" [12]. We have previously reported on students' underuse of help in the Cognitive Tutor environment [6]. The current result suggests that the underuse of help may be a result of students' over-confidence in their ability.

The main limitation of this study is its scope. Not enough data is available on students' SA patterns with more extensive practice, as well as students' spontaneous SA behavior and their domain-level learning gains in the Cognitive Tutor environment.

This work makes several contributions. First, we describe a unique system for tutoring SA skills. The system uses established principles of metacognitive tutoring [15] to help students learn how to evaluate their ability. Second, we demonstrate how classroom research using interactive learning environments can be used to inform our understanding of metacognitive processes. For example, log-file analysis was used to decompose factors that affect students' SA, and to better understand the relationship between domain knowledge and accuracy of SA. Most importantly, this work demonstrates the ability of intelligent tutoring systems to help students improve their metacognitive skills in a manner that transfers to unsupported tasks within the tutoring system.

References

1. Azevedo, R., Moos, D., Greene, J., Winters, F., Cromley, J.: Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia?. Ed. Tech. Res. And. Dev. 56 (1), 45-72 (2008).
2. Aleven, V., Koedinger, K.R.: An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. Cog. Sci. 26 (2), 147-179 (2002)
3. Bunt, A., Conati, C., & Muldner, K.: Scaffolding self-explanation to improve learning in exploratory learning environments. In proceedings of the International Conference on Intelligent Tutoring Systems, 656-67. Maceio, Brazil (2004)
4. Roll, I., Aleven, V., Koedinger, K.R.: The Invention Lab: Using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. Aleven V., Kay J., Mostow J.(Eds.) Proceedings of the International Conference on Intelligent Tutoring Systems, 115-24. Springer Verlag. Berlin (2010)
5. Biswas, G., Roscoe, R., Jeong, H., Sucler, B.: Promoting self-regulated learning skills in agent-based learning environments. In Proceedings of the 17th International Conference on Computers in Education [CDROM] (2009)
6. Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R.: Improving Students' Help-Seeking Skills Using Metacognitive Feedback In An Intelligent Tutoring System. Learning and Instruction 21, 267-280 (2011)
7. Dunning, D., Heath, C., Suls, J.M.: Flawed Self Assessment. Psychological Science In The Public Interest. 5 (3), 69-106 (2004)
8. El Saadawi, G.M., Azevedo, R., Castine, M., Payne, V., Medvedeva, O., Tseytlin, E., Legowski, E., Jukic, D., Crowley, R.S.: Factors affecting feeling-of-knowing in a medical intelligent tutoring system: the role of immediate feedback as a metacognitive scaffold. Adv. Health. Sci. Educ. Theory. Pract., (2009)
9. Nelson-Le Gall, S., Kratzer, L., Jones, E., DeCooke, P.: Children's self-assessment of performance and task-related help seeking. J. Exp. Child. Psychol. 49, 245-263 (1990)
10. Gama, C.: Metacognition in interactive learning environments: the reflection assistant model. In: Proceedings of the International Conference on Intelligent Tutoring Systems, 668-77. Springer-Verlag, Berlin (2004)
11. Bull, s., Kay, j.: Student Models that Invite the Learner In: The SMILI Open Learner Modelling Framework. Int. J. of Art. Int. in Ed.. 17 (2), 89-120 (2007)
12. Kruger, J., Dunning, D.: Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. J of Personality and Soc. Psychol, 77 (6), 1121-1134 (1999)
13. Reder, L.M., Ritter, F.E.: What Determines Initial Feeling of Knowing? Familiarity With Question Terms. J. Exp. Psychol. Learn. Mem. Cogn. (3), 435 - 451 (1992)
14. VanLehn, K. The behavior of tutoring systems. Int. J of AI in Ed, 16(3), 227-265.
15. Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R.: Designing for metacognition - applying cognitive tutor principles to the tutoring of help seeking. Metacognition and Learning. 2 (2), 125-140 (2007)
16. Mathan, S. A. & Koedinger, K. R.: Fostering the Intelligent Novice: Learning from errors with metacognitive tutoring. Ed, Psychol. 40 (4), 257-265. (2005)
17. Aleven, V., McLaren, B.M., Sewall, J., Koedinger, K.R.: A new paradigm for intelligent tutoring systems: Example-tracing tutors. Int. J. of AI In ED. 19 (2), 105-154 (2008)
18. Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R.: Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. Luckin R., Koedinger K.R., Greer J.(Eds.) Proceedings of the International Conference on Artificial Intelligence in Education, 203-10. IOS Press. Amsterdam (2007)