Long, Y., Aleven, V. (2013) Skill Diaries: Improve Student Learning in an Intelligent Tutoring System with Periodic Self-Assessment. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), Proceedings of the 16th International Conference on Artificial Intelligence in Education, AIED 2013, (pp 249-258), New York: Springer. doi: 10.1007/978-3-642-39112-5_26

Skill Diaries: Improve Student Learning in an Intelligent Tutoring System with Periodic Self-Assessment

Yanjin Long and Vincent Aleven

Human Computer Interaction Institute, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
ylong@cs.cmu.edu, aleven@cs.cmu.edu

Abstract. According to Self-Regulated Learning theories, self-assessment by students can facilitate in-depth reflection and help direct effective self-regulated learning. Yet, not much work has investigated the relation between students' self-assessment and learning outcomes in Intelligent Tutoring Systems (ITSs). This paper investigates this relation with classrooms using the Geometry Cognitive Tutor. We designed a paper-based skill diary that helps students take advantage of the tutor's Open Learner Model to self-assess their problem-solving skills periodically, and investigated whether it can support students' self-assessment and learning. In an experiment with 122 high school students, students in the experimental group were prompted periodically to fill out the skill diaries, whereas the control group answered general questions that did not involve active self-assessment. The experimental group performed better on the post-test, and the skill diaries helped lower-performing students to significantly improve their learning outcomes and self-assessment accuracy. This work is among the first empirical studies that successfully establish the beneficial role of self-assessment in students' learning of problem-solving tasks in ITSs.

Keywords: Skill diaries, problem solving, periodic self-assessment, intelligent tutoring system, open Learner model

1 Introduction

Researchers of Intelligent Tutoring Systems (ITSs) have been studying how to enhance students' metacognition in order to support their domain-content learning in ITSs, focusing for example on goal setting, self-explanation, help-seeking, gaming the system, and error correction [6, 11]. Some studies demonstrate that metacognitive support in ITSs can significantly improve students' domain level learning outcomes [6]. However, there has not been much work that investigates students' self-assessment in ITSs, which is also a critical metacognitive skill. Self-assessment refers to students' ability to evaluate their learning status (how well they are learning/have learned). It is thought to be important in two ways. First, the process of self-assessing may help students reflect on their learning, which might result in improved learning outcomes [5]. Second, according to theories of self-regulated learning, accurate self-assessment can help students make good future learning plans [13].

Empirical studies from cognitive and educational psychology have demonstrated a correlation between accurate self-assessment and good learning outcomes. That is, students who assess their own learning more accurately tend to have better learning outcomes [2]. Further, Thiede and colleagues [10] found that improved self-assessment can lead to better re-study choices during learning. However, previous work mainly studied the relationship in the context of memory tests or reading comprehension, whereas ITS researchers tend to focus on problem solving. The nature of self-assessment of problem-solving abilities may well be different from simple memory tests or reading comprehension.

Although not much work has been conducted, some ITS researchers have found interesting and promising results regarding self-assessment. Roll et al. [8] designed a self-assessment tutor that scaffolded students' self-assessment at the start of each section of the tutor curriculum. They found that this tutor improved students' self-assessment on better-mastered problems and that students were able to transfer improved self-assessment in other tutor units [8]. However, this study did not look at whether the self-assessment tutor also enhanced students' domain level learning [8]. Feyzi-Behnagh, Khezri and Azevedo [4] found that by providing metacognitive prompts and feedback, students' self-assessment accuracy improved as well as their learning efficiency (but not the learning effectiveness) when learning with an ITS. Therefore, in spite of these promising initial results it is still an open question how an ITS can support accurate self-assessment in a way that improves robust learning.

A number of researchers have recognized the potential of inspectable Open Learner Models (OLMs) to support students' self-assessment and learning outcomes [1]. However, the promise is not always met. For example, Hartley and Mitrovic [5] compared students' learning gains with or without access to an OLM, but they did not find a significant difference between the two conditions. They only found the less able students' performance improved significantly from pre- to post-test in both conditions [5]. In a previous interview study related to the Geometry Cognitive Tutor [7], a widely-used type of ITS [3], we found that students inspect the tutor's OLM (the "Skillometer") frequently, underlining its potential to support students' self-assessment. We also found, however, that they do not actively use it to reflect or self-assess and that students' self-assessment appears not to be significantly influenced by the Skillometer [7]. Thus, simply presenting an inspectable OLM by itself may not be an effective way to support self-assessment, and additional scaffolding may be necessary. It is an open question what form of scaffolding might be most effective and how interactive it will need to be. White and Frederiksen [12] found that paper-based periodic reflective activities can enhance students' learning significantly. Hence a periodic paper-based method that scaffolds students' use of the Skillometer to help with self-assessing may be similarly effective in an ITS. Therefore, as a first step towards enhancing the Skillometer with self-assessment support, we created a structured, paper skill diary that prompts students to keep track of their skill growth (aided by the Skillometer) while they are learning with a Cognitive Tutor. We conducted a classroom study to test the hypothesis that periodically using the skill diaries can enhance both students' self-assessment accuracy and their learning of math problem-solving skills with the Geometry Cognitive Tutor.

2 Methods

2.1 Participants, Experimental Design, and Procedure

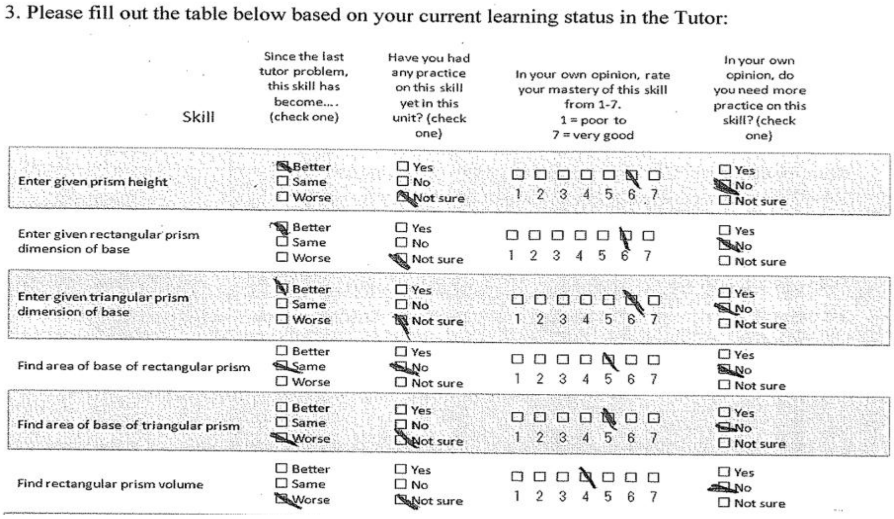
We conducted the study in a local public high school in Pittsburgh in which the Geometry Cognitive Tutor is used as part of the mathematics instruction. A total of 122 students participated and were randomly assigned to two conditions (experimental vs. control). The experimental group periodically filled out skill diaries during their work with the Cognitive Tutor, while the control group periodically answered general questions about the tutor unit they were working on with a control diary. The students came from two math teachers' 6 Geometry Cognitive Tutor classes. For a total of three class periods (around 45 minutes per period), the students covered four sections of the Cognitive Tutor that dealt with volume and surface area of prisms and spheres.

The two groups followed the same procedure: they were first given a pre-test, learned with the Cognitive Tutor for three class periods over consecutive school days, and were then given a post-test following the last tutor class. After the pre-test, the two versions of the diaries (described below) were handed out to the students. During each of the three Cognitive Tutor class periods, the teachers prompted the students to stop twice to fill out the skill/control diaries.

The pre-tests and post-tests were isomorphic and incorporated structurally equivalent Cognitive Tutor problems and transfer problems. There were two parts on both tests. In part I, the to-be-solved problems were shown to the students, while they only needed to rate "How confident are you that you can solve this problem" on a 7-point Likert scale. In part II, students actually solved the problems.

2.2 The Skill Diary and Control Diary

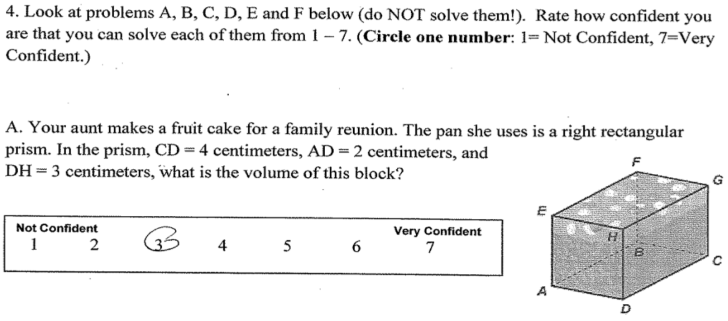
We designed the skill diary to facilitate students' self-assessment both on the skill level and the problem level. There were two kinds of entries in the skill diary: regular entries and end of the day entries. During the three class periods, students were prompted by the teachers to stop and fill out one regular entry twice per class period, and filled out an end of the day entry at the end of each class period. For each of the regular entries, there were three major self-assessment tasks. Firstly, students needed to copy their skill bars from the Skillometer. Secondly, they answered a series of questions in regard to each of the skills listed in the Skillometer, such as "Since the last Tutor problem, this skill has become better/worse/the same?", "Have you had any practice on this skill yet in this unit? Yes/No/Not Sure", and "In your own opinion, do you need more practice on this skill? Yes/No/Not Sure" (Figure 1 shows a filled out diary page for this task). These questions aimed to facilitate students' active self-assessment with the help of the Skillometer. Thirdly, students were asked to rate several specific tutor problems regarding how confident they are in solving these problems based on a 7-point Likert scale (Figure 2 shows an example). The confidence rating on tutor problems was included to enhance students' self-assessment and reflection on the specific problems they encounter in the tutor. It took students about 5 minutes to fill out one regular entry. At the end of each class period, students needed to fill out an end of the day entry that asked them to reflect on their overall learning for that day.



3. Please fill out the table below based on your current learning status in the Tutor:

Skill	Since the last tutor problem, this skill has become.... (check one)	Have you had any practice on this skill yet in this unit? (check one)	In your own opinion, rate your mastery of this skill from 1-7. 1 = poor to 7 = very good	In your own opinion, do you need more practice on this skill? (check one)
Enter given prism height	☒ Better ☐ Same ☐ Worse	☐ Yes ☐ No ☒ Not sure	☐ ☐ ☐ ☐ ☐ ☒ ☐ 1 2 3 4 5 6 7	☐ Yes ☒ No ☐ Not sure
Enter given rectangular prism dimension of base	☒ Better ☐ Same ☐ Worse	☐ Yes ☐ No ☒ Not sure	☐ ☐ ☐ ☐ ☐ ☒ ☐ 1 2 3 4 5 6 7	☐ Yes ☒ No ☐ Not sure
Enter given triangular prism dimension of base	☒ Better ☐ Same ☐ Worse	☐ Yes ☐ No ☒ Not sure	☐ ☐ ☐ ☐ ☐ ☒ ☐ 1 2 3 4 5 6 7	☐ Yes ☒ No ☐ Not sure
Find area of base of rectangular prism	☐ Better ☒ Same ☐ Worse	☐ Yes ☒ No ☐ Not sure	☐ ☐ ☐ ☐ ☒ ☐ ☐ 1 2 3 4 5 6 7	☐ Yes ☒ No ☐ Not sure
Find area of base of triangular prism	☐ Better ☐ Same ☒ Worse	☐ Yes ☐ No ☒ Not sure	☐ ☐ ☐ ☐ ☒ ☐ ☐ 1 2 3 4 5 6 7	☐ Yes ☒ No ☐ Not sure
Find rectangular prism volume	☐ Better ☐ Same ☒ Worse	☐ Yes ☐ No ☒ Not sure	☐ ☐ ☐ ☒ ☐ ☐ ☐ 1 2 3 4 5 6 7	☐ Yes ☒ No ☐ Not sure

Fig. 1. Self-assessment on the skill level in the regular entry of the skill diary



4. Look at problems A, B, C, D, E and F below (do NOT solve them!). Rate how confident you are that you can solve each of them from 1 – 7. (**Circle one number**: 1= Not Confident, 7=Very Confident.)

A. Your aunt makes a fruit cake for a family reunion. The pan she uses is a right rectangular prism. In the prism, CD = 4 centimeters, AD = 2 centimeters, and DH = 3 centimeters, what is the volume of this block?

Not Confident						Very Confident
1	2	[✓ 3]	4	5	6	7

Fig. 2. Self-assessment on the problem level in the regular entry of the skill diary

We also designed a control diary that simply asked students general questions about their learning process, such as “have you seen this problem so far in this unit?” These questions were designed to *not* spur or facilitate active self-assessment. The layouts and structure of the skill diary and control diary were designed as similar as possible to avoid introducing confounding factors between groups.

3 Results

We gathered valid data for 47 students in the control group and 48 in the experimental group. We analyzed students’ pre-test and post-test performance, Cognitive Tutor log data and self-assessment accuracy. We report partial η^2 for effect sizes of main effects and interactions. An effect size partial η^2 of .01 corresponds to a small effect, .06 to a medium effect, and .14 to a large effect (Cohen's guidelines for effect sizes).

3.1 Test Performance on Pre and Post Tests

First, we analyzed whether there were significant learning gains from pre-test to post-test. There were 7 problems on the pre-test and 10 problems on the post-test. The pre- and post-tests shared 5 items that were in the same format but had differing numbers. Students' answers were graded from 0 to 1, with partial credit where appropriate.

To assess the students' improvement from pre-test to post-test, we compared their performance on the shared items. Overall, both groups improved significantly from pre- to post-test (repeated measures ANOVA, F (1, 93) = 13.103, p = .000, η^2 = .123) on the whole test. The group differences were not significant on the pre-test or the post-test. We then divided the test items into two categories: reproduction (isomorphic to the problems in the tutor) and transfer problems. We found that the experimental group did significantly better than the control group on the reproduction problems on the post-test (F (1, 93) = 3.861, p = .052, η^2 = .040), but we found no significant difference between two groups on transfer problems (F (1, 93) = .056, p = .814, η^2 = .001)[1]. In sum, scaffolding students' self-assessment with offline skill diaries lead to better learning, although not better transfer of knowledge.

Table 1. Means and SDs for Reproduction and Transfer Problems (Shared Items)

	Pre-Test (Reproduction)	Post-Test (Reproduction)	Pre-Test (Transfer)	Post-Test (Transfer)
Experimental Group	0.545 (.340)	0.620 (.292)	0.499(.217)	0.579(.263)
Control Group	0.456 (.444)	0.494 (.333)	0.464 (.218)	0.567 (.238)

We also investigated the effectiveness of the skill diary for different ability groups. We expected the skill diaries to be especially effective for the lower-performing group, with respect to both domain level learning and self-assessment accuracy. This expectation was based on prior results by Hartley and Mitrovic [5], who found that an inspectable OLM had a stronger influence on the learning of lower-performing students. We used the median pre-test score (.557) to divide the sample into a lower-performing group with 47 students (average pre-test score: .362) and a higher-performing group with 48 students (average pre-test score: .707). Table 2 shows the higher and lower performing students' performance on pre- and post-test. For the lower-performing students, the difference between conditions on post-test reproduction problems was significant (F(1, 44) = 4.586, p = .038, η^2 = .094; pre-test reproduc-

[1] Although we did not find a significant group effect on the pre-test, when we used the pre-test scores as co-variate, the difference between two groups on reproduction problems was on the borderline of significance (F (1, 92) = 2.747, p = .101, η^2 = .029), suggesting that part of the difference between the two conditions might be accounted for by pre-test differences.

tion problem score was used as co-variate), whereas no significant condition effect was found within the higher-performing group. No significant condition effects were found for transfer problems within the two ability groups either.

Table 2. Means and SDs for Reproduction Problems by Ability Groups

	Pre-Test (Experimental)	Pre-Test (Control)	Post-Test (Experimental)	Post-Test (Control)
Lower-Performing Group	0.346 (.451)	0.163 (.350)	0.527 (.468)	0.300(.390)
Higher-Performing Group	0.744 (.409)	0.738 (.752)	0.713 (.382)	0.679 (.414)

3.2 Process Measures from Cognitive Tutor Log Data

Next, we investigated how the scaffolded self-assessment activities (i.e., the skill diaries) may have influenced students' learning processes within the tutor. Metacognitive processes themselves are unobservable, which is why we looked in the log data for learning behaviors that may be strongly related. Specifically, we looked at: 1) the number of tutor hints students requested; 2) the time students spent on each hint they received from the tutor; 3) the number of incorrect attempts in the tutor; 4) the average assistance score ((hints + incorrect attempts)/total number of steps) in the tutor and 5) the average time students spent on each step. Repeated measures ANOVAs were used with these five process measures from the four tutor sections. The condition (experimental or control) was used as the independent variable. Previous Cognitive Tutor learning data indicated that the four targeted sections vary significantly in their difficulty levels. We found that:

1) The control group asked for significantly more hints per step than the experimental group. The main effect of condition was significant (F (1, 93) = 4.762, p = .032, η^2 = .049).

2) The experimental group spent significantly more time per hint received. The main effect of condition was significant (F (1, 138) = 5.265, p = .023, η^2 = .037).

3) The control group made more incorrect attempts per step. The main effect of condition was marginally significant (F (1, 93) = 3.006, p = .086, η^2 = .031).

4) The control group had a significantly higher assistance score. The main effect of condition was significant (F (1, 93) = 5.388, p = .022, η^2 = .055). The control group also needed more assistance (compared to the experimental group) in the more difficult sections. The interaction between condition and tutor sections was marginally significant (F (3, 279) = 2.281, p = .080, η^2 = .024).

5) The control group spent more time (compared to the experimental group) to finish each step in the more difficult sections. The interaction between condition and tutor sections was significant (F (3, 279) = 2.624, p = .051, η^2 = .027).

Correlations between Process Measures and Test Performance. We calculated the Pearson correlations between these measures and students' test scores. These correlations can help us further interpret whether the differences between conditions on the process measures suggest more effective learning for the experimental condition. As shown in Table 3, the number of hints, number of incorrect attempts and average assistance score are highly correlated with students' pre- and post-test scores, and the negative correlations mean that students with better test performance needed less help and made fewer errors in the tutor. Additionally, the time spent on each hint is significantly correlated with post-test scores. The positive correlations between this process measure and test scores point out that students who have better test performance spent more time studying each hint they received.

Table 3. Correlations between Process Measures and Test Performance

	Number of Hints	Time Spent on Each Hint	Number of Incorrect Attempts	Average Assistance Score	Time Spent on Each Step
Pre-Test	-.558 (.000)**	.199 (.087)	-.350 (.000)**	-.519 (.000)**	-.188 (.067)
Post-Test	-.474 (.000)**	.336 (.003)**	-.317 (.002)**	-.466 (.000)**	-.199 (.053)

** indicates significant level <.01

3.3 Accuracy of Self-Assessment

We also looked at whether the skill diaries influenced the accuracy with which students assessed their own problem-solving ability. Schraw [9] summarized two traditional approaches to measure students' self-assessment accuracy: the relative accuracy and absolute accuracy. For relative accuracy, Gamma and Pearson correlations have been widely used by researchers. For absolute accuracy, Schraw introduced the following formula:

$$\text{Absolute Accuracy Index} = \frac{1}{N}\sum_{i=1}^{N}(c_i - p_i)^2 \qquad (1)$$

where "N" represents the number of tasks, "c" stands for students' confidence ratings on their ability to finish the task while "p" represents their actual performance on that task. The index thus measures the discrepancy between self-assessed and actual performance. The higher the absolute accuracy index, the worse students' self-assessment is. In this paper we only report the results of absolute accuracy. The Gamma correlations were also calculated and led to similar conclusions.

Table 4 shows the absolute accuracy of self-assessment for both conditions. Repeated measures ANOVAs (with the condition as the independent variable) revealed that both groups improved significantly from pre- to post-tests on accuracy of self-assessment (main effect of test time (pre/post): F (1, 93) = 4.369, p = .039, η^2 = .045). The interaction between condition and test time was not significant (F (1, 93) = .023, p = .881, η^2= .000), nor was the main effect of condition (F (1, 93) = .798, p = .374, η^2 = .009).

Table 4. Means and SDs of the Two Groups' Absolute Accuracy of Self-Assessment

	Pre-Test	Post-Test
Experimental Group	0.290 (.133)	0.253 (.128)
Control Group	0.270 (.137)	0.238 (.108)

We compared the self-assessment accuracy of higher- and lower-performing students, given previous work that suggests that better students tend to be more accurate in their self-assessment [2]. As shown in Table 5, on both tests the higher-performing group had a lower absolute self-assessment accuracy score, which indicates more accurate self-assessment of their learning. One-way ANOVAs show that the differences between higher- and lower-performing students on pre-test and post-test were both significant (F (1, 94) = 18.699, *p* = .000, η^2 = .167 and F (1, 94) = 10.064, *p* = .002, η^2 = .098). This finding is aligned with previous literature [2].

Table 5. Means and SDs of Absolute Accuracy of Self-Assessment by Ability Groups

	Pre-Test	Post-Test
Lower-Performing Group	0.336 (.153)	0.283 (.109)
Higher-Performing Group	0.226 (.086)	0.209 (.117)

Next we looked at the higher- and lower-performing groups separately. Within the lower-performing group, paired T-Tests revealed that students in the experimental condition improved significantly with respect to self-assessment accuracy from pre-test to post-test (t(23) = 2.257, *p* = .034), whereas students in the control group did not. Within the higher-performing group, there were no reliable differences between the conditions.

4 Discussion, Conclusion and Future Work

Theories of self-regulated learning emphasize the importance of accurate self-assessment, but little is known about how self-assessment of problem-solving skills (as opposed to memory or reading comprehension) relates to learning, whether and how supporting self-assessment might lead to better skill acquisition, and what kind of support is most effective. The learner modeling capabilities of ITS would seem to provide unique advantages not shared with other learning technologies, as argued in the introduction, but to what extent is this promise met? We investigated whether skill diaries, designed to help students take advantage of an OLM to self-assess periodically, had beneficial effects with respect to learning outcomes and self-assessment accuracy. The results show that students who learned with skill diaries performed better on post-test reproduction problems, compared to control group students, especially the lower-performing students. The results support the hypothesis that periodic self-assessment scaffolded by an OLM can significantly enhance students' learning. This work is among the first empirical studies that successfully establish the beneficial role of self-assessment in students' learning of problem-solving tasks in ITSs.

To better understand how skill diaries might enhance learning, we analyzed tutor log data to study and compare the learning behaviors of students with and without the skill diaries. This analysis revealed differences in learning behaviors between the conditions. Students who learned with skill diaries needed fewer hints but spent more time on the hints they requested, which pointed to more appropriate use of help from the tutor. Correlation analysis also revealed that the time students spent on each hint positively correlate with their test scores. Furthermore, in more difficult sections of the tutor, the control group spent more time on each step and had a higher average assistance score. Both the time per step and average assistance score correlate negatively with students' test scores, which suggests that the experimental group students learned more effectively and efficiently in harder sections.

The results from log data suggest how the use of a skill diary might enhance students' learning outcomes. Firstly, when prompted to copy their skill bars and answer specific self-assessment questions both on the skill and problem levels, students might be more likely to notice skills that they have not yet mastered, as well as problems they are not yet good at. They might then reflect on the errors they made on these skills and problems, as well as on how they corrected them with help from the tutor or their teachers. Such reflection and self-assessment may be more rare without skill diaries. Secondly, based on theories of self-regulated learning [13], self-assessment can help students to direct attention and effort to address the content that they have not yet mastered. Despite the structured nature of Cognitive Tutors, students can regulate their learning in that they decide when to receive help messages from the tutor. Therefore, when students went back to the tutor after filling out the diary, with their self-assessment in mind, they might use the tutor's hints more deliberately, which could help them master the not-yet-mastered skills. Thirdly, the diaries explicitly directed students' attention to the change of their skill bars, which might help them be more alert and motivated to stay focused on their learning. The fewer incorrect attempts in the tutor may have provided evidence for this change in students' learning behaviors. In the future, we may conduct think-alouds and interviews to further investigate the mechanisms of how the skill diary or periodic self-assessment works to enhance students' learning outcomes.

We also found significant improvement on the accuracy of self-assessment for lower-performing students who used the skill diaries. Previous studies [2] have documented students' overconfidence when self-assessing their learning status, which was more severe for the lower performing students. Skill diaries may have broken the illusion of mastery for the lower-performing students during the learning process, so they could form a more objective view of their learning.

We did not find significant benefits for higher-performing students, with respect to both the learning outcomes and self-assessment accuracy. It is possible that the higher-performing students already possess good self-assessment, so there is not much room for improvement. But it will still be worth investigating in the future why the intervention was more helpful for lower-performing students, and how we can support all students' self-assessment and learning outcomes effectively.

To sum up, both test results and log data from the present study help to empirically establish the beneficial role of self-assessment in learning of problem-solving tasks in ITSs. Although theories of self-regulated learning have emphasized the critical role of self-assessment in learning, our study is among the first rigorous classroom studies which have successfully illustrated the benefits of periodic self-assessment for problem-solving tasks in ITSs. The critical features of the skill diary, namely, prompting students' self-assessment periodically both on the skill level and problem level, can be transferred to build online tools integrated with the OLMs that support students' self-assessment and metacognition in ITSs.

Acknowledgments. We thank Gail Kusbit and Tristan Nixon for their kind help with this work. We would also like to thank the participating teachers and students. This work is funded by an NSF grant to the Pittsburgh Science of Learning Center (NSF Award SBE0354420).

References

1. Bull, S. Supporting Learning with Open Learner Models, *Proceedings of 4th Hellenic Conference with International Participation: Information and Communication Technologies in Education*, Athens, Greece. Keynote, (2004)
2. Chi, M., Bassok, M., Lewis, M., Reimann, P., Glaser, R. Self-explanations: How students study and use examples in problem solving. *Cognitive Science*, 13, 145-182, (1989)
3. Corbett, A., McLaughlin, M., Scarpinatto, K. Modeling Student Knowledge: Cognitive Tutors in High School & College. *User Modeling and User-Adapted Interaction*, 10, 81-108, (2000)
4. Feyzi-behnagh, R., Khezri, Z., Azevedo, R. An Investigation of Accuracy of Metacognitive Judgments during Learning with an Intelligent Multi-Agent Hypermedia Environment, *The annual meeting of the Cognitive Science Society*, Cognitive Science Society, 96-101, (2011)
5. Hartley, D., & Mitrovic, A. Supporting Learning by Opening the Student Model. In S. A. Cerri, G. Gouardères & F. Paraguaçu (Eds.) *Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, Berlin: Springer Verlag, 453-462, (2002)
6. Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. In vivo Experiments on Whether Supporting Metacognition in Intelligent Tutoring Systems Yields Robust Learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education*, 897–964, New York, NY: Routledge, (2009)
7. Long, Y., Aleven, V. Students' Understanding of Their Student Model. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.) *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, 179-186, Berlin: Springer Verlag, (2011)
8. Roll, I., Aleven, V., McLaren, B., & Koedinger, K. R. Metacognitive practice makes perfect: Improving students' self-assessment skills with an intelligent tutoring system. In G. Biswas, S. Bull, J. Kay, & T. Mitrovic (Eds.) *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, Berlin: Springer Verlag, 288-295, (2011)
9. Schraw, G. A Conceptual Analysis of Five Measures of Metacognitive Monitoring. *Metacognition and Learning*, 4, 33–45, (2009)
10. Thiede, K.W., Anderson, M.C.M., Therriault, D. Accuracy of Metacognitive Monitoring Affects Learning of Texts. *Journal of Educational Psychology*, 95, 66–73, (2003)
11. Weerasinghe, A., Azevedo, R., Roll, I., du Boulay, B. *The Proceedings of the Fourth Workshop on Self-Regulated Learning in Educational Technology*, 11th International Conference on Intelligent Tutoring Systems, (2012)
12. White, B. C., & Frederiksen, J. Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students, *Cognition and Instruction*, 16, 39-66, (1998)
13. Winne, P. H., & Hadwin, A. F. Studying as Self-Regulated Learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice*, Hillsdale, NJ: Erlbaum, 279-306, (1998)