# Feedback for Metacognitive Support in Learning by Teaching Environments

**Jason Tan and Gautam Biswas (jason.tan, gautam.biswas@vanderbilt.edu)**
Department of EECS &ISIS, Vanderbilt University
Nashville, TN 37235 USA

**Daniel L. Schwartz (daniel.schwartz@stanford.edu)**
School of Education, Stanford University
Stanford, CA 94305 USA

## Abstract

Past research on feedback in computer-based learning environments has shown that corrective feedback helps immediate learning, whereas guided and metacognitive feedback help in gaining deep understanding and developing the ability to transfer knowledge. Feedback becomes important in discovery learning environments, where novice students are often overwhelmed by the cognitive load associated with learning and organizing new knowledge while monitoring their own learning progress. We focus on feedback mechanisms in teachable agent systems to help improve students' abilities to monitor their agent's knowledge, and, in the process their own learning and understanding. Our studies demonstrate the effectiveness of guided metacognitive feedback in preparing students for future learning.

## Introduction

Metacognition has been identified as a critical process that supports student learning and problem solving (Bransford, Brown & Cocking, 2000). Brown (1987) describes two component processes: (i) the ability to monitor one's cognitive activities, and (ii) the ability to take appropriate regulatory steps when a problem has been detected. These steps can include internal regulation (e.g., slow down when reading hard material) and external action (e.g., consult learning resources). Both abilities increase with maturation (Flavell, 1987), but appropriate educational opportunities can propel metacognitive development and improve subsequent learning. We focus on developing learning environments that provide metacognitive support and examining whether metacognitive interventions improve students' subsequent abilities to learn.

Self-monitoring (cf. to self-explanation (Chi, et al., 1994)) is a key metacognitive strategy that supports learning with understanding, and the ability to apply the learnt knowledge to problem solving tasks. However, self-monitoring is itself a complex cognitive task. In the context of problem solving, it requires two simultaneous coordinated "processes": one that develops a sequence of steps to solve the problem, and a second that evaluates the correctness and efficiency of the problem solving process. Analyzing discrepancies and making corrections adds further complexity to the self-monitoring task. For example, when solving math problems, one ideally runs a systematic procedure that computes a precise answer, and a second process that does a quick and approximate analysis to estimate an approximate answer. Answers that are far out of alignment should trigger debugging activities to resolve the discrepancy.

There is evidence that helping children learn to monitor others problem solving can, in turn, help them monitor their own problem solving and learning. For example, Palinscar and Brown (1986) found that a strong emphasis on monitoring improved students' reading and learning abilities. Ideally, monitoring someone else's work should prompt the children to run their own process to generate a solution, and then compare it against the other person's solution. This means that students need not do both processes simultaneously. This reduces cognitive load, develops the awareness and capacity to compare solutions, and with time makes it easier to turn this capacity inward.

Our proposed research leverages this hypothesis using Teachable Agents (TAs), which are software environments where students teach a computer agent using well-structured visual representations. The TA reasons with the facts and relations it has been taught to answer questions and solve problems. Using their agent's performance as a motivation, students work to remediate the agent's knowledge, and learn better on their own. One of our TAs, called Betty's Brain, has been successfully used to teach river ecosystems in 5th grade science classrooms (Biswas, et al. 2005a, 2005b).

An important property of Betty's Brain is that students monitor how Betty answers questions and can correct her (and themselves) when she makes mistakes. However, 5th grade students, who are both domain novices and novices in teaching practices, often do not possess the necessary monitoring skills, and they often fail to analyze relevant pointers to errors and omissions in their knowledge. This has led us to develop metacognitive cues as explicit feedback mechanisms within the TA environment to help students develop the monitoring abilities. This paper discusses the effectiveness of these feedback mechanisms in aiding the students monitoring and learning tasks. The results of an experimental study in a 5th grade science classroom are discussed in terms of the students' immediate learning abilities and their preparation for future learning (Schwartz and Martin, 2004).

## Betty's Brain

Betty, shown in Fig. 1, is taught using a concept map representation. Students teach her about entities, such as fish and algae, and their relations, (e.g., fish consume dissolved oxygen, algae replenish it) in river ecosystems. Once taught, Betty uses qualitative reasoning methods to reason through chains of links (Biswas, et al., 2005a), which helps her answer questions, such as "*if macroinvertebrates increase what happens to bacteria?*" Learning by teaching is implemented as three primary components: (i) *teach* Betty by constructing a concept map, (ii) *query* Betty with your own questions to see how much she has understood, and (iii) *quiz* Betty with provided tests to see how well she does on questions you may not have considered. These activities are usually embedded within a larger narrative (e.g., teach Betty so she can pass a test to join a science club).

When asked, Betty explains her answers using text, speech and animation. Students reflect on Betty's answers and revise their own knowledge as they make changes to the concept maps to teach Betty better. Our work has demonstrated that one of the primary benefits of learning by teaching a TA is the need to structure knowledge in a compact and communicable format so that the student-teacher may develop important explanatory structures for the domain. The fact that TAs have independent performance and can show their reasoning based on how they have been taught also helps students (and teachers) assess their teaching (and by implication, learning). This should provide metacognitive and self-assessment opportunities for students that can lead to superior learning and transfer.

To help novice students with their learning and teaching tasks, we built in additional resources into the environment: (i) domain resources organized as searchable hypertext so students can look up information as they teach Betty, (ii) a concept map tutorial that provides students information on causal structures, and how to reason with these structures, and (iii) a Mentor agent, Mr. Davis, who provides on demand feedback about learning, teaching, and domain knowledge ("Ask Mr. Davis"). The Mentor also provides feedback immediately after Betty takes a quiz. We have designed two different versions of the Mentor agent. One agent gives corrective feedback to students and the other agent provides guided feedback in the form of metacognitive strategies. We discuss this in more detail in the next section.

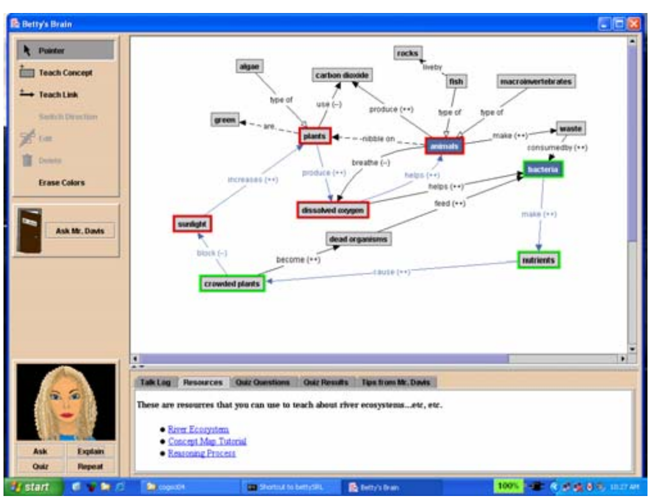


**Figure 1: Betty's Brain – Interface**

### Metacognitive Support and Preparation for Future Learning (the PFL study)

A past study conducted in 5th grade science classrooms showed evidence that learning-by-teaching with metacognitive support for self-regulated learning helped students develop better learning strategies, and prepared them better for future learning on related topics, even when this learning happens outside of the support provided by the TA environment (Biswas, et al., 2005a). Students were divided into groups to work on three versions of the system: (i) Intelligent Tutoring System (ITS), where the Mentor asked students to create a concept map that would correctly answer a set of test questions, (ii) Learning by Teaching (LBT), where students taught Betty to help her pass a test to become a member of the school Science club, and (iii) Self-Regulated Learning (SRL), where students taught Betty for the same reason as the LBT group, but the Betty persona incorporated metacognitive learning strategies (Pintrich and DeGroot, 1990; Zimmerman, 1989). All three groups had access to identical resources on river ecosystems, the same quiz questions, and similar access to the query feature and the Mentor agent.

The differences in performance for the three groups in the main study were not significant (we studied the quality of the concept maps students generated and the quiz scores). However, we expected the SRL students would do better than the others, in a *preparation for future learning* task, where they had to learn a new topic on their own. All students were asked to construct a concept map to answer questions about a topic they had not studied before, the land-based nitrogen cycle. They had access to resources but there was no feedback from Betty or the Mentor. The SRL group created maps with more concepts and links than the ITS and LBT groups. The effects of teaching self-regulation strategies had an impact on the students' abilities to learn a new domain (Biswas, et al., 2005a, Schwartz, et al., in press). These results encouraged us to study metacognitive feedback and its effects on student learning in a more systematic way.

## Feedback and Student Learning

A number of studies have demonstrated the effectiveness of feedback in computer-based learning environments (for a review see (Azevedo and Bernard, 1995)). Two forms of content feedback, i.e., *directed* or *corrective* feedback and *guided metacognitive* feedback have been used extensively to assist student learning. Moreno (2004) used feedback for decreasing the cognitive load of novice students in discovery-based multi-media environments. Her study compared a guided learning environment, where an agent provided explanatory feedback with a directed learning environment, where the agent provided corrective feedback. Her guided discovery hypothesis centered on the belief that learning occurs when students actively construct a coherent knowledge representation by meaningful interactions with resource materials, converting the information extracted into representations, and integrating new information into existing representations. Typically, discovery learning results in high cognitive load for students with low prior knowledge, making it hard for them to learn. Her studies showed that the guided feedback group found the instructional material easier to follow, made significantly fewer errors on post test questions, and was much better at transfer tasks that involved novel situations than the directed feedback group.

Aleven and Koedinger (2002) performed studies on students' help seeking behavior with a Cognitive Tutor for Geometry. The system provided on demand help at multiple levels of detail, starting with general strategies relevant to a problem solving step to specific hints, which explicitly outlined the correct solution for that step. In initial studies, the researchers found that most students quickly clicked down to the most detailed corrective hints and ignored the general strategy and theoretical hints. It was not clear that this feedback improved overall student learning. In later work Aleven, et al. (2004) incorporated self explanation, where the students were required to explain their problem solving steps. In addition to corrective error feedback, the system provided guided self explanation hints centered on general strategies for finding knowledge related to the current problem. Students showed deeper learning when the tutor required them to explain their steps. Students in the explanation condition spent more time on the system than students who were not required to provide self-explanations, but they needed fewer problems to achieve predetermined mastery levels for skills.

These studies support our early findings with the Betty's Brain system. Corrective feedback aids students with immediate problem solving tasks, but it is unclear that it produces deep learning and self monitoring skills, especially when the supporting feedback is removed. We observed this in the above study, where the ITS and LBT students, who received corrective feedback, did better in the main study quiz than the SRL students, who mostly received guided feedback (Fig. 2). As discussed earlier, when it came to the transfer test, students in the ITS condition were frustrated because of they were not prepared to learn on their own. This led a number of ITS students to give up after the first transfer study session (Biswas, et al., 2005b). The SRL students, who had received guided feedback and metacognitive support in the main study developed better learning and monitoring strategies that they were able to apply in the transfer study (Butler and Winne, 1995; Biswas et al., 2005b).

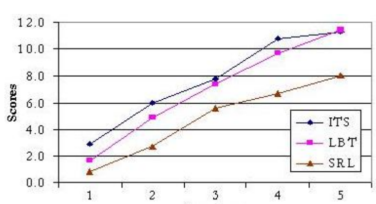


**Figure 2: Average Quiz scores at the end of each main study session (max score = 15)**

## The Different Types of Feedback in Betty's Brain

In order to further study the role of metacognitive feedback, we performed a new study which examines the effects of corrective versus guided feedback in developing students self-monitoring and self-regulation skills. We started with the existing LBT and SRL versions of the system.

The LBT system provides corrective feedback. After every quiz attempt, the Mentor compares Betty's answers to that generated by the expert concept map[1] and informs her (and the student) about right and wrong answers. For every incorrect answer, the Mentor first checks if the concepts in the quiz question appear in the student's map. If they do not, the Mentor suggests that the student study these concepts in the resources. Otherwise, the Mentor compares the student's map with the expert's to look for the first (i) missing expert concept, (ii) missing expert link, and (iii) mismatch between expert and student map link, in that order in the causal chain, to generate the appropriate feedback content. Like the Cognitive Tutor (Aleven and Koedinger, 2002) the Mentor provides hints that range from general (e.g., "read about algae and dissolved oxygen") to specific ("you are missing a link between algae and dissolved oxygen in your concept map"). We believe that this form of corrective feedback focuses students on the task of getting their quiz questions right as opposed to trying to understand domain content. This results in a trial and error behavior that we have labeled as the *quiz/edit/quiz pattern* (Biswas, et al., 2005b).

Our SRL system feedback is designed to teach students a set of comprehensive skills. This involves setting goals for learning new materials, developing strategies that lead to effective problem solving, monitoring one's learning progress, and then revising one's knowledge, beliefs, and strategies to overcome errors and to assimilate new content. Betty's persona in the SRL version incorporates self-regulation and metacognitive strategies. For example, when the student is building the concept map, she occasionally responds by demonstrating reasoning through chains of events. She may query the user, and sometimes remark (right or wrong) that the answer she is deriving does not seem to make sense. The idea behind these spontaneous prompts is to get the student to reflect on their own knowledge, and like a good teacher check on their tutee's learning progress. At times, Betty asks the students to query her and ensure that she reasons correctly with the concept map. Also, Betty refuses to take a quiz, saying she has not been taught enough, or that the student has not tested her learning by asking her queries.

The Mentor and Betty's interactions are driven by an activity-tracking system that derives *patterns of behavior* from the students' activities on the system and Betty's performance on the quizzes. We believe that Betty's and the Mentor's feedback in the SRL condition helps students develop better self-regulation and self-monitoring strategies that carry over to subsequent learning tasks even in environments where the feedback is not present. Self-monitoring abilities will manifest in setting goals that are linked to gaining knowledge on quiz questions that Betty has not answered correctly. As a result, students will consult relevant resources periodically to gather information, which they will use to update and correct their concept maps. Betty's insistence on being queried to check whether she had learnt the material correctly before students ask her to take a quiz, will promote the use of the query and explanation feature as self-monitoring tools. These behaviors developed from feedback received in the main study should carry over to future learn-

[1] The student never sees the expert map.

ing tasks in environments where the same feedback may not be present.

In addition to comparing direct corrective versus guided metacognitive feedback, we were also very interested in determining how the guided feedback content may affect the student's learning behaviors. We created two versions of the SRL system that take their cues from the same set of patterns but provide very different kinds of feedback. The SRL-Cognitive (SRL-C) feedback is content directed, and students are given hints that help them apply metacognitive strategies to improve their learning, monitoring, and debugging tasks. The SRL-Affective (SRL-A) feedback uses the same pattern cues, but Betty's responses are emotional rather than content-oriented (see **Table 1**).

**Table 1: Behavior patterns and SRL-C and SRL-A responses**

| Pattern | Cognitive Response | Affective Response |
|---|---|---|
| If after four questions, Betty has not been queried on an unlinked concept | Excuse me. You taught me a concept, but didn't teach me any relationships between it and other concepts. Please teach me more, and ask me questions to make sure I understand | Hey, I'm confused and I don't understand what you taught me. Please teach me more, and ask me some questions. |
| If quiz and causal query but no update | Hey, you haven't taught me anything new since my last quiz. My score will surely be the same. Teach me something, and ask me some questions to make sure I understand, before you send me to take another quiz | Hey! You're making me do really hard things and I don't like it. |
| If no resource access and no improvement on previous quiz score | Excuse me. I like what you are teaching me, but it may not help me pass the quiz. I would like to be better prepared when I take it again. Could you check the resources and teach me about what you find there? Thanks. | Excuse me, but that quiz is very difficult. I really don't want to take it now. Can we do something else? |

## Experimental Study and Results

This new study, conducted in two 5th grade classrooms, was designed to compare the effects of the different types of feedback. 39 students from the two classrooms were divided into the three groups (SRL-C, SRL-A, and LBT) using a stratified sampling method based on standard achievement scores in mathematics and language. The students worked with Betty's Brain for seven 45-minute sessions. Their goal was to successfully teach Betty about river ecosystems and get her to pass three quizzes (answer all questions correctly). Approximately 10 weeks later, the students were given the transfer test, where they taught Betty about the land-based nitrogen cycle. The students worked for three sessions and this permitted us to determine which group was better prepared to learn in situations where scaffolds and feedback from their previous environments were removed. We believed that students previously in the SRL-C condition would demonstrate the best performance for future learning, and students in LBT condition with directed feedback would perform better than the students in the SRL-A condition, who received no useful feedback.

### Experimental Results

Both performance and behavioral data were analyzed in this study. Students' activities on the TA systems in the main and transfer study were recorded in log files, along with Betty's and Mr. Davis' feedback. The students' concept maps were also saved at the end of each session. In evaluating the students' concept maps we marked all concepts and links that appeared in the expert map as well as other concepts and links that were considered to be correct in describing the domain as valid concepts and links.

**Table 2: Main Study: Valid Concepts and Links in Students Concept Maps by session (concepts on top, links below)**

| Session number | LBT Mean (sd) | SRL-A Mean (sd) | SRL-C Mean (sd) |
|---|---|---|---|
| 1 | 2.64 (1.96) | 2.77 (1.36) | 3.31 (1.25) |
| | 0.82 (1.40) | 1.23 (1.24) | 1.00 (0.82) |
| 2 | 8.55 (3.05) | 8.69 (3.90) | 9.23 (3.06) |
| | 3.55 (2.95) | 3.85 (2.82) | 5.08 (2.63) |
| 3 | 11.27 (4.96) | 12.62 (5.71) | 13.46 (5.24) |
| | 6.64 (5.22) | 7.38 (6.06) | 10.15 (5.83) |
| 4 | 12.55 (6.06) | 14.69 (8.58) | 17.54 (8.02) |
| | 9.55 (7.22) | 9.38 (9.30) | 13.54 (7.84) |
| 5 | 13.73 (5.20) | 16.15 (9.86) | 18.54 (8.93) |
| | 11.55 (7.75) | 11.62 (11.30) | 14.77 (9.12) |
| 6 | 14.55 (6.61) | 17.62 (10.00) | 19.69 (9.28) |
| | 13.09 (8.37) | 13.54 (11.73) | 15.77 (9.70) |
| 7 | 14.27 (6.21) | 18.00 (9.88) | 21.69 (10.96) |
| | 13.36 (8.02) | 14.92 (12.59) | 18.15 (10.09) |

At the end of the main study, the LBT group had more links from the expert map (7.2) than the two SRL groups (SRL-A=5.6 and SRL-C=5.3), but these differences were not statistically significant. However, as Table 2 shows, the SRL-C group had more valid concepts and links in their maps than the LBT and the SRL-A groups. The numbers for SRL-A group were about the same as the LBT group. ANOVA indicated all of the groups showed significant improvements in their map quality over time, but the between group differences were not significant. As discussed earlier, we believe that corrective feedback led the LBT students to primarily focus on the concepts and links required to get the quiz answers right. The metacognitive feedback for the SRL-C group directed the students to focus on acquiring relevant knowledge from the resources and organize it into the concept map structure somewhat independent of the specific quiz questions. We believed this would promote better abilities to learn on one's own.

To check this, we looked at the transfer test concept maps. The number of expert concepts for all three groups was about the same (differences between groups were not statistically significant). However, the SRL-C group had the largest number of valid concepts, and this difference was statistically significant (Table 3). Overall, students did not perform well in the quiz,[2] but the SRL-C group had more valid links than the other two groups (Table 3). In other words, they were better at extracting information from the text resources and creating valid concept map structures, which implied better learning performance in the new domain.

[2] We determined that students did not make do well with the nitrogen cycle because we did not give them sufficient time to learn the difficult concepts and relations from the resources.

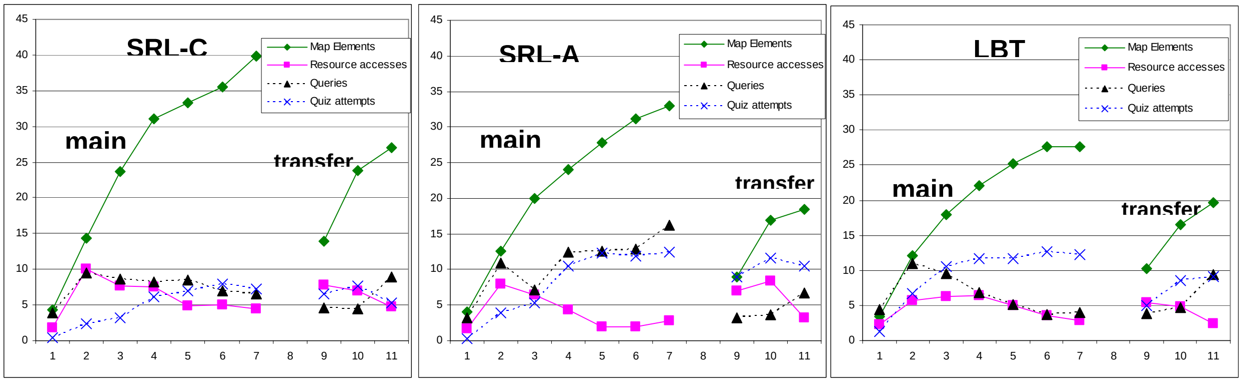


**Figure 3: Number of Map elements (valid links + valid concepts), Queries, Quiz Attempts, and Resource Accesses by session in the Main and Transfer Study for all three groups.**

This demonstrates the effectiveness of metacognitive, guided feedback in preparation for future learning.

**Table 3: Transfer Study: Number of Valid Concepts and Links in Students' Final Map**

| Transfer test | SRL-C Mean (sd) | SRL-A Mean (sd) | LBT Mean (sd) |
|---|---|---|---|
| Valid concepts | 14.69[ab] (5.5) | 10.23 (4.9) | 10.27 (5.6) |
| Valid links | 10.85 (7.6) | 8.5 (6.5) | 9.3 (6.9) |

[a] Significantly greater than SRL-A, $p < .05$;
[b] Significantly greater than LBT, $p < .05$.

In addition to evaluating students' concept maps, we also monitored their behaviors in the main study and the transfer test. Analyzing the student log files revealed differences between the three groups that can be attributed to the differences in the feedback they received. We focus on quiz attempts, queries asked, and resource accesses as they demonstrate the students' abilities to monitor their learning and seek new information. The average counts for the three variables are plotted by session and group for the main study and the transfer study in Fig. 3.

Beginning from session 4 in the main study the SRL-C group showed a balanced behavior pattern in quiz, query, and resource use. The SRL-A group showed a large drop in resource accesses but the number of queries and quiz attempts remained high. For the LBT group, both resource accesses and queries decreased after session 4, but the number of quiz attempts remained high. Asking Betty to take a quiz allowed the students to monitor Betty's progress and their own teaching and learning. However, making Betty repeatedly take quizzes without systematic attempts to gain new information, update the concept map, and debug the map by generating queries, showed that the LBT and SRL-A students resorted to trial-and-error behaviors that we previously characterized as the *quiz/edit/quiz* pattern. In the SRL system, Betty refuses to take the quiz a second time unless the student updates the concept map or queries her to see if she has understood what she has been taught (**Table 1**). This feedback helped the SRL-C group students to develop better learning and teaching behaviors, and this is seen in Fig. 4, where the number of quizzes refused by Betty for this group is much smaller than for the SRL-A group, where the number of quiz refusals was high throughout the main study (Fig. 4). Table 4 shows the sessions where the differences in quiz refusals between the two SRL groups were significant. Although Betty refused to take a quiz in exactly the same situations for both SRL groups, the SRL-C got over the trial-and-error strategy and adopted systematic learning methods. Interestingly, the SRL group's systematic learning behavior continued in the transfer test even when the feedback was no longer present.

**Table 4: Quiz attempts refused**

| Main study | SRL-C Mean (sd) | SRL-A Mean (sd) |
|---|---|---|
| Quiz attempts refused (sess. 2) | 1.77 (1.4)[a] | 3.54 (2.2) |
| Quiz attempts refused (sess. 3) | 1.92 (1.8)[a] | 4.15 (3.3) |
| Quiz attempts refused (sess. 7) | 4.08 (3.9)[a] | 9.33 (6.8) |

[a] Significantly less than SRL-A, $p < .05$

The SRL-A group asked many more queries of Betty than the SRL-C and LBT groups in sessions 6 and 7 (Table 5). Whereas the SRL-C group realized the role of the query feature in monitoring Betty's and their own knowledge, it is clear that the SRL-A group generated queries only to get Betty to take the quiz. This is further supported by the fact that they had many more quiz attempts refused as reported above. In the transfer test, the differences in the number of queries asked by group are not significant. This indicates that the SRL-A group did not learn to use queries for debugging their map.

**Table 5: Queries per session**

| Main study | SRL-C Mean (sd) | SRL-A Mean (sd) | LBT Mean (sd) |
|---|---|---|---|
| Queries (sess. 6) | 7.00 (6.0) | 12.92 (9.1)[ab] | 3.73 (3.8) |
| Queries (sess. 7) | 6.62 (6.7) | 16.17 (13.7)[ab] | 4.00 (4.6) |

[a] Significantly more than SRL-C, $p < .05$
[b] Significantly more than LBT, $p < .05$

Further support of these claims can be made by correlational analysis of the main study data. Only the SRL-C group demonstrates a significant correlation between quiz attempts and number of added valid links per session. The negative correlation value indicates that the students who were successful in adding valid links did not repeatedly use the quiz feature (Pearson r=-0.247, n=91, p=0.009), and this behavior persisted in the transfer test. A similar analysis showed that both the SRL-C and LBT groups exhibited a positive correlation between queries asked and number of added valid links per session (Pearson r for SRL-C: r=0.196, n=91, p=0.031; for LBT: r=0.226, n=77, p=0.024), but this did not hold for the SRL-A group.

The continued balanced behavior pattern in the transfer study combined with the fact that the SRL-C students had

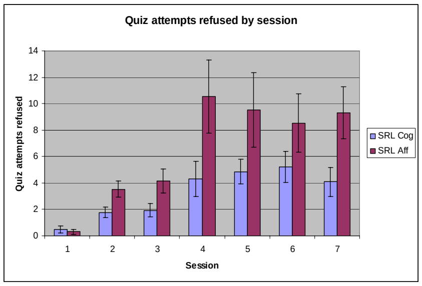


**Figure 4: Main study: Number of quizzes refused**

more valid concepts and links, shows that they used the quiz, query, and resources effectively even when the scaffolds and feedback were removed (Fig. 3). The large discrepancy between queries and resource accesses was repeated for the SRL-A, and the LBT group was somewhere in between.

## CONCLUSIONS

The results from this study demonstrate the performance and behavioral differences in learning that can be associated with the different types of feedback provided by our TA system. Although corrective feedback may allow the student to achieve immediate goals set by the learning environment quickly, like earlier work (Aleven and Koedinger, 2002; Moreno, 2004), we have demonstrated that guided metacognitive feedback better prepares the student for future learning tasks even in situations where the metacognitive support is removed. However, guided feedback with metacognitive cues but no content information does not help novice learners with low prior knowledge. Students have to be taught and given enough opportunities to practice metacognitive strategies in socially engaging and relevant ways.

The differences between the SRL-C, SRL-A and LBT groups indicate that the type of feedback received has a significant effect on learning outcomes. This was illustrated in the transfer study, when the students from the three different conditions were asked to learn a new domain in an environment with very little scaffolding and feedback. Analysis of student log files showed that the LBT group that received corrective feedback followed a *quiz/edit/quiz* behavior pattern that was mostly focused on getting quiz questions right, whereas the SRL-C group that received guided feedback triggered by metacognitive cues showed a balanced use of the query, quiz, and resource features indicating the use of self-monitoring strategies when learning new domain content. These behaviors were repeated in the transfer test when all support was removed indicating that the SRL-C students were better prepared for future learning.

## Acknowledgements

This work has been supported by a NSF ROLE Award #0231771. The help provided by other members of the Teachable Agents group, Kadira Belynne, Bilikiss Adebiyi, Krittaya Leelawong, Nancy Vye, Joan Davis, and John Bransford are gratefully acknowledged.

## References

Aleven V. and Koedinger, K. R. (2002) An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science,* 26, 147-179.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2004) Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. *Proc.7th Intl. Conf. on Intelligent Tutoring Systems,* 227-239.

Azevedo, R. and R.M. Bernard. (1995) A meta-analysis of the effects of feedback on computer-based instruction. *Jour. of Educational Computing Research*, 13, 111-127.

Biswas, G., D. Schwartz, K. Leelawong, N. Vye, and TAG-V. (2005a) Learning by Teaching: A New Agent Paradigm for Educational Software, *Applied Artificial Intelligence*, special issue on Educational Agents, 19, 363-392.

Biswas, G., et al. (2005b) Case Studies in learning by Teaching: Behavioral Differences in Directed versus Guided Learning. *27th Annual Meeting of the Cognitive Science Society,* Stresa, Italy, 828-833.

Bransford, J.D., A.L. Brown, and R.R. Cocking, eds. (2000) *How People Learn*. National Academy Press, Washington, D.C.

Brown, A.L. (1987) Metacognition, Executive Control, Self-Regulation and other more Mysterious Mechanisms. *Metacognition, Motivation, and Understanding* Weinert and Kluwe (eds.), LEA, Hillsdale, NJ, 65-116.

Chi, M. T. H., et al. (1994) Eliciting self explanations. *Cognitive Science,* 18, 439-477.

Flavell, J.H. (1987) Speculations about the Nature and Development of Metacognition, *Metacognition, Motivation, and Understanding,* Weinert and Kluwe (eds.), LEA, Hillsdale, NJ, 65-116.

Moreno, R. (2004) Decreasing Cognitive Load for Novice Students: Effects of Explanatory versus Corrective Feedback in Discovery-Based Multimedia. *Instructional Science*, 32, 99-113.

Palincsar, A. S., and Brown, A. L. (1984) Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. *Cognition and Instruction*, 1, 117-175.

Pintrich, P. R. and E. V. DeGroot. (1990) Motivational and self-regulated learning components of classroom academic performance, *Journal of Educational Psychology,* 82, 33-40.

Schwartz, D. L. and Martin, T. (2004) Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. *Cognition & Instruction*, 22, 129-184.

Schwartz, D.L., K.B. Pilner, G. Biswas, K. Leelawong, and J.P. Davis. (in press) Animation of Thought: Interactivity in the Teachable Agent Paradigm. *Learning with Animation: Research and Implications for Design*, Lowe and Schnotz (eds.), Cambridge Univ. Press.

Zimmerman, B. J. (1989) A Social Cognitive View of Self-Regulated Academic Learning, *Journal of Educational Psychology*, 81, 329-339.