Promoting Motivation and Self-Regulated Learning Skills through Social Interactions in Agent-based Learning Environments

Gautam Biswas, Hogyeong Jeong, Rod Roscoe, and Brian Sulcer

Department of EECS/ISIS
Box 351829, Sta B
Vanderbilt University, Nashville TN 37235

Abstract

We have developed computer environments that support learning by teaching and the use of self regulated learning (SRL) skills through interactions with virtual agents. More specifically, students teach a computer agent, Betty, and can monitor her progress by asking her questions and getting her to take quizzes. The system provides SRL support via dialog-embedded prompts by Betty, the teachable agent, and Mr. Davis, the mentor agent. Our primary goals have been to support learning in complex science domains and facilitate development of metacognitive skills. More recently, we have also employed sequence analysis schemes and hidden Markov model (HMM) methods for assigning context to and deriving aggregated student behavior sequences from activity data. These techniques allow us to go beyond analyses of individual behaviors, instead examining how these behaviors cohere in larger patterns. We discuss the information derived from these models, and draw inferences on students' use of self-regulated learning strategies.

Introduction

We have developed computer-based learning environments that use the learning by teaching paradigm to help middle school students develop higher-order cognitive skills when learning in science and math domains (Biswas, et al., 2005; Leelawong and Biswas, 2008). To teach, one must gain a good understanding of the domain knowledge and then structure the knowledge in a form that they can present to others (Bargh and Schul, 1980). Preparing to teach is a self-directed and open-ended activity where one explores, integrates, and structures knowledge first for oneself, and then for others. Biswas, Schwartz, & Bransford (2001) have reported that students preparing to teach felt that the responsibility to teach encouraged them to gain deeper understanding of the materials. Beyond preparing to teach, actual teaching taps into the three critical aspects of learning interactions – *structuring, taking responsibility, and reflecting*. With respect to structuring, teaching peers gives students opportunities to organize their knowledge and articulate it via explanations, which facilitates self-monitoring and revision of that knowledge. Moreover, interactions with the pupil (e.g., questions) can prompt additional reflective knowledge-building for the teacher (Roscoe and Chi, 2007). For taking responsibility, teaching is frequently open-ended and self-directed, and teachers need to take the responsibility of deciding which content is most relevant (Artzt and Armor-Thomas, 1999). Finally, for reflection, effective teaching requires the explicit monitoring of how well ideas are understood and used. Studies have shown that tutors and teachers often reflect on their interactions with students during and after the teaching process in order to better prepare for future learning sessions (Chi, et al., 2001).

We have designed a teachable agent (TA) system called Betty's Brain, where students teach a computer agent using

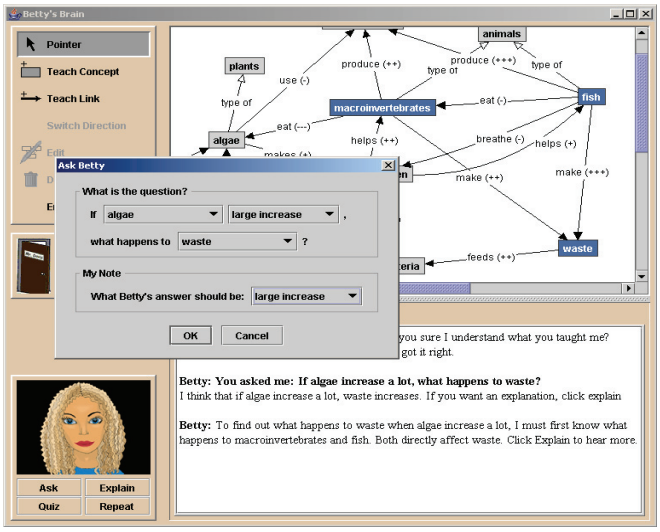


Figure 0: Betty's Brain System with Query Window

a well-structured visual representation (Biswas, et al, 2005; Leelawong and Biswas, 2008). Using their agent's performance (which is a function of how well it is taught) as a motivation, students learn for themselves so that they can remediate the agent's knowledge, and, in this process, they learn better on their own. For this reason our *learning-by-teaching* environments are well-suited to helping students become more knowledgeable of and responsible for their own cognition and reasoning. As a result, the students are likely to develop problem solving and monitoring skills that go beyond the learning of specific domain content; rather they provide the much larger framework that guide students on *how to learn* and how to *prepare for future learning* (Schwartz and Martin, 2004). We have hypothesized that working with Teachable Agents helps students better understand domain knowledge, and engage in a variety of productive learning strategies that promote organizing and reasoning with this knowledge. Furthermore, the activities involved in the teaching process helps the students monitor their own learning as they teach their agent.

This paper discusses the results of a study we conducted in 5th grade science classrooms, where students taught their agent about entities and their relationships in a river ecosystem. One of our goals was to determine if teaching an agent produced better learning performance than students who learnt for themselves. A second goal was to check if metacognitive prompts by the TA and Mentor agents helped the student develop metacognitive strategies that they applied to their learning. Given these goals, this paper focuses on analyses of students' behaviors as they teach Betty by creating their concept maps. Such analyses are important because they shed light on students' underlying learning processes, and what kind of strategies they are bringing to this task (Roscoe and Chi, 2007).

Learning by Teaching: The Betty's Brain System

The Betty's Brain system is illustrated in Figure 1. The teaching process is implemented as three primary activities: (i) *teach:* Students explicitly teach Betty using a concept map representation (Novak, 1998) that includes concept names, which appear as boxes, and links between concepts that appear as arrows. The links can be of two types: (a) descriptive, and (b) causal. Students teach Betty new concepts and links using the Teach Concept and Teach Link buttons. They can also delete and modify their concepts and links using the Delete and Edit buttons; (ii) *query:* Students use a template to ask Betty questions and find out how she answers them based on what she has been taught; and (iii) *quiz:* Students observe Betty's performance on a set of predefined questions that are presumably assigned by a Mentor agent. The questions are typically organized into sets of quizzes.

Once taught, Betty uses qualitative reasoning methods to reason through chains of links (Forbus, 1984; Biswas, et al., 2005) to answer questions, and, if asked, explain her reasoning using text and animation schemes. Betty also provides feedback that reflects the students' teaching behaviors. The goal is to get the students to adopt more metacognitive strategies in their learning tasks (Tan, Biswas, and Schwartz, 2006, Wagster, et al., 2007; Jeong and Biswas, 2008). Students reflect on Betty's answers and her explanations, and revise their own knowledge as they make changes to the concept maps to teach Betty better. Further details of the Betty's Brain system and earlier experiments that we have conducted with this system are summarized in (Biswas, et al., 2005; Leelawong and Biswas, 2008). Next we discuss the self regulated learning support provided to students as they learn about river ecosystems.

Metacognitive Support in Betty's Brain

Cognitive science researchers have established that metacognition and self-regulation are important components in developing effective learners in the classroom and beyond (Bransford, Brown, and Cocking, 2000; Zimmerman, 2001). Pintrich (2002) differentiates between two aspects of metacognition for learners: (i) *metacognitive knowledge* that includes knowledge of general strategies and when they apply, as well as knowledge of one's own abilities, and (ii) *metacognitive control and self regulatory processes* that learners use to monitor and regulate their cognition and learning. We believe the TA environments when combined with adequate scaffolding and feedback can provide appropriate educational opportunities for students to develop both metacognitive knowledge and control, and thereby, improve their subsequent learning.

We have adopted a self-regulated learning (SRL) framework that describes a set of comprehensive skills that start with setting goals for learning new materials and applying them to problem solving tasks, deliberating about strategies to enable this learning, monitoring one's learning progress, and then revising one's knowledge, beliefs, and strategies as new materials and strategies are learnt (Azevedo, 2005; Schraw, Kauffman, & Lehman, 2002; Winne and Hadwin, 2008; Zimmerman, 2001). In conjunction with these higher level cognitive activities, social interactions and motivation also play an important role in the self-regulation process (Zimmerman, 20010). We believe that two interacting factors of our TA implementations are particularly supportive of self regulation. The first is the visual *shared representation* that the students use to teach their agents. The second factor, *shared responsibility,* targets the positive effects of social interactions to learning. This manifests in the form of a joint effort where the student has the responsibility for teaching the TA (the TA knows no more and no less than what the student teaches it), whereas the TA assumes responsibility for answering questions and taking tests.

Betty's persona in the SRL version incorporates metacognitive awareness that she conveys to the students at appropriate times to help the student develop and apply monitoring and self regulation strategies (Wagster, et al, 2007; Schwartz, et al., 2009). We have identified a number of recurrent interactive action sequences, where metacognitive feedback might be useful for students. When the system detects such patterns, Betty provides suggestions on strategies the students may employ to improve their own

understanding of the subject matter. For example, at times, when students add concepts and links, Betty remarks that the relation between two connected concepts does not make sense to her. At other times, Betty spontaneously responds by demonstrating reasoning with "chains of links." In some ways, this helps the students become more familiar with the reasoning processes that Betty uses to derive her answers, especially indirect effects, something that the students may not be very familiar with. Even more, this may be a cue for students to reflect on what they have taught, check if Betty derives the right answers, and go read the resources again, if they are not sure. Some of the other triggering patterns related to goal setting, self assessment and monitoring are illustrated in Table 1. Betty's responses combine motivational and self-regulation cues, whereas the Mentor, Mr. Davis' responses focus primarily on self regulation strategies. Mr. Davis' responses are linked to the students' activity patterns. For example, if the student gets Betty to repeatedly take the quiz and not read the resources, the Mentor agent reminds Betty and her student teacher the importance of reading resources, and checking ones understanding after learning (teaching) new material.

Table 1. Some Interactive Action Patterns with Betty's and the Mentor's responses

Regulation Goal	Pattern Description	Betty Response	Mr. Davis' Response
MONITORING THROUGH EXPLANATION	**Multiple requests for Betty to give an answer but no request for explanation**	Let's see, you have asked me a lot of questions, but you have not asked for my explanations lately. Please make me explain my answers so you will know if I really understand.	Without asking Betty to explain her answers, you may not know whether she really understands the chain of events that you have been trying to teach her. Click on the Explain button to see if she explains her answer correctly.
SELF-ASSESSMENT	**Repeated quiz request but no updates have been made to the map.**	Are you sure I understand what you taught me? Please ask me some questions to make sure I got it right. I won't take the quiz otherwise. Thanks for teaching me about rivers!	You have not taught Betty anything new. Please, spend some time teaching her new links and concepts and make sure she understands by asking her questions. Then she can take the quiz again. If you need help learning new things, check the resources.

Experimental Design

Our participants were 56 students in two 5th grade science classrooms, taught by the same teacher. Students were assigned to one of three conditions using stratified random assignment based on standardized test scores. All students created concept maps on river ecosystem concepts and causal relations over seven 45-minute sessions. However, two of the groups (i) the learning by teaching (LBT) group, and (ii) a self-regulated learning by teaching (SRL) group were told they were creating their map to teach Betty so that she could pass a test on her own later. As they taught Betty, they could ask her to answer queries, and take quizzes. Our third group, the intelligent coaching system (ICS) group, was told to create the map to learn for themselves. They could also query their map, and submit the map for quizzes, but in this case, the Mentor, Mr. Davis, answered their questions, or told them how well they had fared in the quiz. The ICS condition represented our control condition. In the SRL version, Betty also generated spontaneous responses that were driven the interactive patterns described in Table 1.

All students took a pre-test before they worked on the system, and a post-test after they had finished their seven sessions. The tests contained a set of multiple choice questions, and some free-response questions (see Biswas and Leelawong, 2008 for details).

Students Learning Performance

We used two measure of learning performance: (i) Pre-post test gains on the multiple choice and free-response questions, and (ii) gain in concept map scores. The gain in concept map scores was calculated as the difference between their final map score and the map score at the end of session 1. Table 2 shows the gains in the score by condition.

Table 2. Pre-post test and concept map score gains

Gain Score	Conditions		
	ICS	LBT	SRL
Multiple Choice	0.4 (2.4)	1.1 (3.1)	0.4 (1.5)
Free Response	1.9 (3.0)	4.3 (3.2)	4.8 (4.7)
Map Concepts	8.1 (2.4)	7.3 (2.7)	10.4 (3.1)
Map Links	12.2 (3.8)	12.7 (5.3)	16.2 (4.4)

It is clear that the two groups that taught Betty (LBT & SRL) outperformed the ICS group on all measures (Free Response, SRL > ICS, $p < 0.1$; Map Concepts, SRL > ICS, $p < 0.05$, SRL > LBT, $p < 0.01$; Map Links, SRL > ICS, $p < 0.05$, SRL > LBT, $p < 0.1$). The SRL condition had better scores than the LBT group (the differences were not significant), implying that the SRL feedback may have helped students in their learning and monitoring tasks.

Table 3. Gain score correlations

Gain Score	Free Response	Map Concepts	Map Links
Multiple Choice	**.16**	**-.07**	**.13**
Free Response	--	**.35**[a]	**.41**[a]
Map Concepts	--	--	**.54**[c]

[a]$p < .05$. [b]$p < .01$. [c]$p < .001$.

Table 3 shows the correlations between the different gain scores. The free response questions, which require students to reason about important concepts, such as interdependence and balance, and also reason in causal chains, e.g., a change in the amount of algae would affect macroinverte-

brates, which would then affect the fish population, showed strong correlations with the map scores (all of the correlations were significant). On the other hand, the multiple choice questions, asked about definition or direct causal relations. Students could often guess the right answer from the choices provided. Therefore, it is not surprising that the multiple choice scores did not correlate well with map scores.

Analyzing Student Activity Patterns

All student activities in the system were captured in log files. We wrote computer programs that coded these activities into five primary categories:

1. Editing; this included activities like adding, deleting, or changing a concept or link to the students map.
2. Ask Query
3. Take Quiz
4. Read Resources
5. Check Explanation.

The program also recorded a number of off-topic activities, e.g., a student adding concepts and links that had nothing to do with ecosystem domain, and then asking Betty a query just to hear her speak. For each relevant activity, the program captured additional information related to the activity. For example, when the student asked a query, we noted the query (i.e., the set of concepts), and Betty's response to the query.

Three levels of analysis were used for studying the students' behavior patterns. At the first level, we looked for correlations between the frequency of students' activities and their learning as measured by gains in free response test scores (Roscoe, et al., 2008). These results are briefly summarized in the next subsection. Our second level of analysis, we looked for related sequences of students' activities, to see if these sequences implied patterns of behavior that could be linked to their use of metacognitive strategies that have been reported in the literature (e.g., Zimmerman (2001); Winne and Hadwin (2008)). We defined two measures: (i) informedness and (ii) diagnosticity that are discussed in a subsequent subsection to characterize students' use of strategies.

Our third level of analysis was more comprehensive. We used statistical learning methods to derive aggregate behavior models in the form, of hidden Markov Models (HMMs) from students' entire activity sequences as they worked on the system. We briefly describe the HMM approach and the results in the third section.

Level 1 Analysis: Frequency of Student Activities. The average number of edit, query, quiz, read, and explain events by condition are listed in Table 4. On the whole, the LBT group performed many more actions than the ICS group ($F_{2,45} = 8.41$, $p < .001$). The other differences were not statistically significant. The LBT group performed more edit actions than the ICS and SRL groups (LBT > ICS, $p < 0.1$, LBT > SRL, $p < 0.1$). The LBT group also requested more quizzes than the SRL group ($p < .05$), but when it came to query and explanation actions, the SRL group had many more than the ICS and LB groups (queries, SRL > ICS, LBT, all $p < .001$, and explanations, SRL > ICS, $p < .001$, SRL > LBT, $p < 0.1$). The LBT group on the other hand had many more Read events (LBT > SRL, $p < .002$, LBT > ICS, $p < .05$).

Table 4. Frequency of different activities by condition

Gain Score	Conditions		
	ICS	LBT	SRL
Total Events	179.6 (44.2)	258.7 (64.2)	217.8 (51.8)
Edit Events	92.6 (26.6)	118.3 (38.9)	92.7 (29.9)
Query Events	12.2 (9.3)	40.4 (20.6)	67.2 (21.3)
Explain Events	2.4 (2.6)	7.1 (8.4)	12.5 (8.0)
Quiz Events	17.5 (8.1)	25.9 (15.5)	15.8 (7.5)
Read Events	33.8 (18.7)	51.9 (29.3)	23.1 (11.9)
Off-Topic Events	*21.1 (35.5)*	*15.2 (28.3)*	*4.5 (7.3)*

Edit events can be linked to map building activities, whereas query, explanation, and quiz events are related to monitoring activities. Quiz events are related to checking the correctness of one's map, for a set of queries (provided by the Mentor), where Query and Explain events may be considered more advanced monitoring activities, since they involve formulating one's own queries, and then checking how an answer was generated by tracing through the map. Overall, the LBT and SRL groups were much more active than the ICS group during their learning tasks, and had significantly less Off-topics behavior events. The LBT group had the most edit, read, and quiz events, which are good learning behaviors, but the SRL group showed more evidence of advanced monitoring activities, i.e. Query and Explain events.

Level 2 Analysis: Quality of Student Activities. Although students had access to the same features (e.g., queries and explanations, etc.), not all of them used these features effectively. For example, whereas one student might use queries to diagnose the effect of revisions to their map (good use) or to identify Betty's incorrect knowledge (good use), other students might ask questions simply to hear Betty speak (bad use) or to simply confirm what they have just taught (unclear use). For this paper, we assume students learning activities or events fall into one of two states: (1) editing or map building and refining state, and (2) diagnosing state. All editing actions put students in the map building state, and all other actions (query, explanation, quiz, and read) actions move students into the diagnosing state. The simplified student activity model is illustrated in Fig. 2.

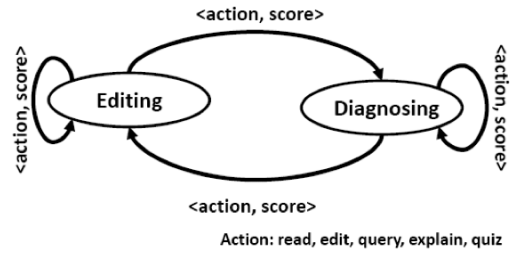


Figure 2: Simplified Strategy Use Model

Each student action is also assigned a score that depends on how closely related it is to previously performed actions. For example, if a student edits a part of the map that is related to the query he just asked, the edit action is related to the previous action, and gets a higher "informedness" score than if the edit action were unrelated to the preceding query. In other words, the informedness scores for edits is based on "how much information about the domain or model was used to make that edit." Scores range from 0-5, and higher scores are better.

Similarly, the diagnosticity scores provide a measure of how "diagnostic" a query, quiz, explanation, or read event is in the way it is used with other activities. In other words, "how much information does the diagnostic action provides students about the effect of map additions and revisions on Betty's "knowledge?" Again scores range from 1-5, and higher scores are better.

Table 5. Informedness and Diagnosticity scores by condition

Gain Score	Conditions		
	ICS	LBT	SRL
Edit Events	2.7 (0.7)	2.9 (0.8)	2.3 (0.6)
Query Events	0.9 (0.7)	1.3 (0.4)	1.4 (0.3)
Explain Events	2.7 (2.9)	4.0 (2.1)	5.3 (1.1)
Quiz Events	3.9 (1.3)	3.0 (1.7)	3.4 (1.5)
Read Events	3.6 (0.8)	3.8 (0.9)	3.2 (1.5)

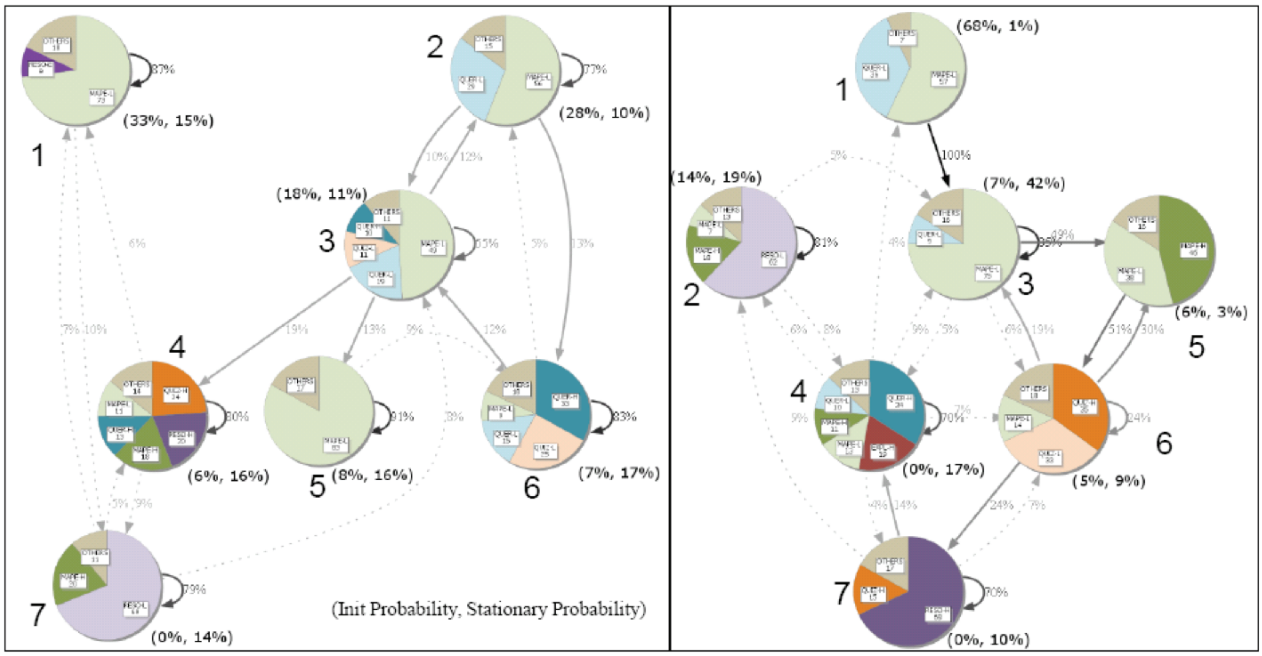
A few significant group differences were observed. The LBT group had a higher read score than the SRL group ($p < 0.1$), whereas the SRL and LBT groups had higher query scores than ICS group (SRL > ICS, $p < .005$, LBT > ICS, $p < .05$), and the SRL group had a higher Explain score in comparison to the ICS group (SRL > ICS, $p < .01$). These results again demonstrate that the SRL group used more advanced probing and checking actions during their learning activities than the other two groups.

Level 3 Analysis: Aggregate Behaviors using hidden Markov Models. A hidden Markov model (HMM) defines dynamic behavior of a system or process as a transition through a sequence of states, with output that corresponds to components of the behavior being described. For example, a student's learning behavior may be described in terms of a read and organize state followed by check and monitor state. In the read and organize state the student behavior is described by activities that include reading resources and building the concept map, and the check and monitor state is defined by activities, such as asking queries, taking the quiz, and re-reading resources. Mathematically, a HMM model is defined by three sets of parameters: initial probability vector π, which indicates the likelihood that a student behaviors will start in a particular state, a transition probability matrix, A, which captures the likelihood that a student will move on from one state to a subsequent state, and output probability matrix, B, which indicates the likelihood of different activities being observed in a particular state (Rabiner, 1989). By representing concise models of student activity patterns, a HMM has the potential of providing us with a global aggregated view of how students approach their learning tasks (Jeong and Biswas, 2008).

We have developed an algorithm that constructs HMMs given a set of activity sequences (Li and Biswas, 2002; Jeong and Biswas, 2008) that uses the Bayesian information criterion (BIC) to trade off simplicity of the model against information provided by the model. In other words, we apply the Occam's razor Principle (simpler is better) to find the model that strikes a balance between high likelihood and low complexity (Li and Biswas, 2002). We extended the processes described in our previous work to generate HMM models in 3 steps: (1) Initialization: A clustering algorithm was applied to initialize the parameters of the HMM model and come up with a definition of the initial model; (2) Model Building: This is the core step, where the parameter optimization algorithms and the Baum-Welch criterion are applied to derive the optimal HMM model for a given set of sequence data; and (3) Model Interpretation: Meaning is assigned to the derived states of the model, and the behavior description is generated in terms of the derived states.

In this part of the study, we divided the students into three groups based on the differences in their pre-post test gains. Specifically, the free response scores in students' posttest were used to group the students into low, mid, and high groups. There were 19 students in the low group (scores ranged from 2 to 6), 14 students in the mid group (scores ranged from 7 to 9), and 16 students in the high group (scores ranged from 10 to 18). The max score on the test was 20 points. Next, the pre-post gains differences were used to divide the students into low_low (LL), low_high (LH), and high_high (HH) groups. The LL group represented students who scored at or below the median in both the pre and the posttest (4 and 8, respectively). The LH group represented students who scored at or below the median in the pretest, scored at or above the median in the posttest, and whose gain in scores was at least 5. The HH group represented students who scored at or above the median in both the pre and the posttest. There were 13 students in the HH group, 12 students in the LH group, and 19 students in the LL group. 5 students could not be put into one of the three groups; their pre and posttest scores were (8,6), (8,5), (6,5), (7,5), and (8,7).

The HMM behavior sequence models for LL and HH groups are shown in Fig. 3. The best fit models for these two models have 7 states each, which are shown as circles in the figure. The arrows between the states indicate possible transitions between states, and the number beside an arrow indicates the likelihood of a transition between states expressed as a percentage. For example, in the LL group model, the likelihood that a student will move to a state 4 behavior from a state 3 behavior is 19%, and the likelihood for moving to a state 5 behavior is 13%. Self-loops indicate the likelihood of continuing to exhibit behaviors in the



Low-Low Group **High-High Group**

Figure 3: HMM Behavior models

same state, for example, state 3 has a self-loop likelihood of 55%. To reduce clutter in the behavior diagrams, we show the more likely transitions as solid lines, and the less likely ones as dotted lines. Using the transition probabilities and the likelihood of starting in a state (shown as the first number in parentheses by a state in the figure) we compute the likelihood of the proportion of the total activities that are associated with a state. For example, the likelihood that a student in the LL group started in behavior state 3 is 18% and 11% of their total activities are associated with that state. The color-coded pie charts within a state represent the different activities the students performed within the state. The informedness and diagnosticity measures were used to categorize an activity as high or low. For example, a query that has a high diagnosticity score is ranked high, Query_H, and a query that has low diagnosticity score is ranked low, Query_L.

To further analyze student learning behaviors and to differentiate among the behaviors of the three groups, we characterize behavior states into: (1) *Uninformed editing* – in this state the students are primarily making uniformed changes to their map, indicating the use of trial and error or guessing strategies. Students may spend some time reading resources or asking queries, but these activities do not relate to their editing activities. It is interesting that all three groups were most likely to start their learning activities in this state; (2) *Reading* –students are primarily engaged in reading the resources in this state. They may combine reading with some editing of their maps. Reading is considered to be a good information seeking behavior; (3) *Checking* – this behavior state combines uninformed editing with asking queries and taking quizzes, but the activities are not related. For example, the query a student asks does not relate to the part of the map the student is editing, or the part of the map the student edits is not related to the quiz questions that had wrong answers. This behavior state is linked to weak monitoring strategies; (4) *Combined Probing and Editing* – this behavior state combines checking, reading, and informed editing during their map building actions; (5) *Probing* – this behavior state combines querying, quizzing, and editing in a way that the map editing is informed by the results of the querying and quizzing functions; and (6) *Advanced Probing* – in this behavior state the combine querying and quizzing with the explanation feature, which provides a trace or the chain of links that were followed to generate an answer to a question. Using this characterization, in the LL behavior model shown in Figure 3, states 1 and 5 represent Uniformed Editing states, state 7 is a reading state, states 2 and 3 represent Checking states, and state 6 is a Probing state, and state 4 represents the Combined Probing and Editing state. In contrast, for the HH group behavior model state 3 is the Uninformed editing state, states 2 and 7 represent the Reading state, state 1 is the Checking state, states 5 and 6 are the Probing states, and state 4 is the Advanced probing state.

Table 6 compares the behaviors across groups. 53% of the LL group's activities are directed to uninformed editing and checking, 14% to reading, and 38% to probing behaviors. In contrast, the numbers for the LH and HH group are 32%, 23%, and 32%, and 42%, 29%, and 9%, respectively. The LH and HH groups exhibit Advanced probing behaviors (use of explanations), which the LL group does not.

Table 6. Behavior States by Group with Likelihood of starting and proportion of activities.

Behaviors	Low-Low (LL)		Low-High (LH)		High-High (HH)	
	Start %	Propor-tion %	Start %	Propor-tion %	Start %	Propor-tion %
Uninformed Editing	41	32	6	17	7	41
Reading	0	14	23	23	14	29
Checking	46	21	63	15	68	1
Probing	7	17	5	32	5	9
Probing & Editing	6	16				
Advanced Probing			3	13	6	20

A clear difference between the HH group and the other two groups is the use of explanations in their learning and map building tasks, and this helped them perform better in their learning tasks. The LL group's inability to learn may be attributed to their inability to apply advanced probing strategies. A second reason may be that they did not spend as much time in reading and learning from the resources as the other two groups did. In summary, the use of reading and Advanced Probing behaviors differentiate the high performers from the low performers.

Table 7. Number of students and completion rates by group.

Performance	ICS		LBT		SRL	
	#	Completed (#,%)	#	Completed (#,%)	#	Completed (#,%)
LL	8	2 (25%)	5	1 (20%)	6	3 (50%)
LH	2	1 (50%)	4	1 (25%)	6	3 (50%)
HH	3	1 (33%)	7	5 (71%)	3	3 (100%)

Table 7 lists the number of students in the LL, LH, and HH groups by condition. We also list the number of students in each group who were able to complete their concept mapping task in the 7 sessions that were allocated to them. It is clear that the SRL condition outperformed the other two groups, indicating that the students who received SRL strategy feedback were able to perform their learning tasks better (an overall 60% completion rate versus 31% for ICS and 44% for LBT). The LBT group performed marginally better than the ICS group. Prior knowledge did impact the ability to complete the learning task for the LBT and SRL groups, but not for the ICS group. It does seem learning by teaching with SRL and metacognitive strategy feedback helped students learn their science content better, but the effects of prior knowledge and effectiveness of the strategy feedback provided will have to be studied further.

Discussion and Conclusions

The Betty's Brain system is designed to leverage the benefits of learning by teaching and causal reasoning to facilitate students' science learning. We have hypothesized that working with Betty is helpful because it supports students' engagement and promotes educationally productive cognitive and metacognitive processes. The results reported here, along with prior research, support this hypothesis. Students who utilized learning by teaching versions of our system (i.e., LBT and SRL versions) constructed better concept maps that captured causal relationships between entities in a river ecosystem than were students who used the non-teaching ICS version of the system. Moreover, students' performance was strongest when we explicitly supported their use of self-regulated learning strategies by having Betty model and prompt for such behaviors.

Although assessments of learning outcomes were in agreement with our hypotheses, it was also critical to explore students' actual behaviors during the teaching and learning process. Did students in the LBT and SRL conditions perform well because they were engaged in productive cognitive and metacognitive behaviors? For this purpose, we developed three levels of analysis: (1) studying the frequency of student activities as they worked with the system, (2) weighing student actions using diagnosticity and informedness measures, and (3) a novel method for examining students' aggregated behaviors using HMMs. Frequency analysis clearly indicated that monitoring behaviors, such as querying and checking explanations correlated strongly with learning. This was substantiated in the level 2 analysis, where the activities were weighed by the context of the surrounding actions. The HMM models provide a more aggregated description of student behaviors. We were able to characterize states of the HMM in terms of SRL strategies that are reported in the literature. Our HMM model derivation process reported in this paper, also gives a better sense of the context in which different actions were used. For example, was an editing action informed by the last query action, or was it unrelated?

In future work, we hope to refine our analyses further to get a better understanding of the different strategies that middle school students employ when learning complex science topics. We will also continue to study the effects of using strategy feedback and guidance to help students become better learners. Another direction that we will pursue is the role of self-efficacy and motivation in learning SRL strategies.

Acknowledgments. This work has been supported by Dept. of ED IES grant #R305H060089, NSF REESE Award #0633856, and NSF IIS Award #0904387.

References

Artzt, A. F., Armour-Thomas, E. (199). Cognitive model for examining teachers' instructional practice in mathematics: A guild for facilitating teacher reflection. In: Educational Studies in Mathematics. vol. 40, issue 30, pp. 211–335.

Azevedo, R. (2005). Using Hypermedia as a Metacognitive Tool for Enhancing Student Learning? The Role of Self-Regulated Learning. In: Educational Psychologist. vol. 40, issue 4, pp. 199--209. Lawrence Erlbaum Associates.

Bargh, J.A., Schul, Y. On the cognitive benefits of teaching. Journal of Educational Psychology, 72 (5). 593-604. (1980)

Biswas, G., Schwartz, D., Bransford, J., et al. (2001). Technology support for complex problem solving: From SAD environments to AI. In: K.D. Forbus and P.J. Feltovich (Eds.), Smart Machines in Education. AAAI Press, pp. 71-98. Menlo Park, CA.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N. (2005). Learning by Teaching: A New Agent Paradigm for Educational Software. In: Applied Artificial Intelligence. AAI, vol. 19, pp. 363-392.

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

Chi, M.T.H., Siler, S. A., Jeong, H., Yamauchi, T., Hausmann, R. G (2001) Learning from Human Tutoring. In: Cognitive Science. vol. 25, issue 4, pp. 471--533. (2001)

Forbus, K., Qualitative Process Theory, Artificial Intelligence, vol. 24, pp. 85-168. (1984)

Jeong, H., Biswas, G. (2008) Mining Student Behavior Models in Learning-by-Teaching Environments, *First International Conference on Educational Data Mining*, Montreal, R. S. Baker, T. Barnes, T., I.E. Beck, (eds.), pp. 127-136, Montreal, Quebec, Canada, June 20-21.

Leelawong, K., Biswas, G. (2008) Designing Learning by Teaching Systems: The Betty's Brain System. In International Journal of Artificial Intelligence in Education.

Li, C., Biswas, G. (2002) A Bayesian Approach for Learning Hidden Markov Models from Data. Special issue on Markov Chain and Hidden Markov Models, Scientific Programming, vol. 10, pp. 201-219.

Novak, J.D (1998) Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations. Lawrence Erlbaum Associations, Mahwah, NJ.

Pintrich, P. (2002) The Role of Metacognitive Knowledge in Learning, Teaching, and Assessing, Theory into Practice: Revising Bloom's Taxonomy, 41(4), 219-225.

Rabiner L. R. (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In: Proc. IEEE. vol. 77, issue 2.

Roscoe, R. D., Chi, M (2007) Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. In: Review of Educational Research. vol. issue. 4, pp. 534-574.

Roscoe, D., Wagster, J., and Biswas, G. (2008) Using Teachable Agent Feedback to Support Effective Learning by Teaching,, Cognitive Science Conference, Washington, DC, July.

Schraw, G., Kauffman, D.F., Lehman, S. (2002) Self-regulated Learning Theory, In L. Nadel (ed.) The Encyclopedia of Cognitive Science, pp. 1063-1073, London: Nature Publishing Company.

Schwartz, D.L., Chase, C., Chin, C., Oppezzo, M., Kwong, H., Okita, S., Biswas, G., Roscoe, R.D., Jeong, H., Wagster, J.D. (2009)). Interactive Metacognition: Monitoring and Regulating a Teachable Agent, *Handbook of Metacognition in Education*, Hacker, D.J., Dunlosky, J., Graesser, A.C., Routledge Press.

Tan, J., Biswas, G., Schwartz, D. (2006) Feedback for Metacognitive Support in Learning by Teaching Environments, The 28th Annual Meeting of the Cognitive Science Society, Vancouver, Canada, pp. 828-833.

Wagster, J., Tan, J., Wu, Y., Biswas, G., Schwartz, D (2007) Do Learning by Teaching Environments with Metacognitive Support Help Students Develop Better Learning Behaviors? In: The 29th Annual Meeting of the Cognitive Science Society, Nashville, Tennessee, pp. 695-700.

Winne, P., Hadwin, A. (2008). The weave of motivation and self-regualted learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regualted learning: Theory, research, and applications* (pp. 297-314). NY: Taylor & Francis.

Zimmerman, B. Theories of self-regulated learning and academic achievement: An overview and analysis. In & B. Zimmerman & D. Schunk (Eds.), Self-regulated learning and academic achievement: Theoretical perspectives (pp. 1-37). Mawah, NJ: Erlbaum. (2001)