
Investigating Self-Regulated Learning in Teachable Agent Environments

29

John S. Kinnebrew, Gautam Biswas,
Brian Sulcer, and Roger S. Taylor

Abstract

We have developed a computer-based learning environment that helps students learn science by constructing causal concept map models. The system builds upon research in learning-by-teaching (LBT) and has students take on the role and responsibilities of being the teacher to a virtual student named Betty. The environment is structured so that successfully instructing their teachable agents requires the students to learn and understand the science topic for themselves. This learning process is supported through the use of adaptive scaffolding provided by feedback from the two agents in the system: the teachable agent, Betty, and a mentor agent, Mr. Davis. For example, if Betty performs poorly on a quiz, she may tell the student that she needs to learn more about the topics on which she is performing poorly. In addition, Mr. Davis may suggest that students ask Betty questions and get her to explain her answers to help them trace the causal reasoning chains in their map and find out where she may be making mistakes. Thus the system is designed to help students develop and refine their own knowledge construction and monitoring strategies as they teach their agent.

This chapter provides an overview of two studies that were conducted in fifth-grade science classrooms. A description of the analysis techniques that we have developed for interpreting students' activities in this learning environment is also provided. More specifically, we discuss the generation of hidden Markov models (HMMs) that capture students' aggregated behavior patterns, which form the basis for analyzing students' metacognitive strategies in the system. Our study results show that students who utilized LBT versions of our system performed better than students who used a non-teaching version of the system. Further, students' performances

J.S. Kinnebrew • G. Biswas (✉) • B.Sulcer • R.S. Taylor
Department of EECS & ISIS, Vanderbilt University,
Nashville, TN 37235, USA
e-mail: gautam.biswas@vanderbilt.edu

were strongest when the system explicitly provided support to help them develop self-regulated learning strategies. To gain further insight into the students' reactions to feedback from the two agents, we present results from a second study that employed a think-aloud protocol. Overall, the results from this study illustrated that students were more receptive to the explicit strategy-oriented feedback from the mentor agent. Interestingly, this study also suggested that students had difficulty in correctly applying Betty's feedback related to metacognitive monitoring activities.

Introduction

Cognitive scientists have established that metacognition and self-regulation are important components for developing effective learning in the classroom and beyond (Bransford, Brown, & Cocking, 2000; Zimmerman, 2001). The framework for self-regulated learning (SRL) originated from the social cognitive theory of learning proposed by Bandura (1997), who postulated that learning is governed by three interacting factors: (1) *personal* (e.g., learners attitudes and beliefs); (2) *behavioral* (e.g., the ability to invoke relevant prior knowledge, the ability to employ appropriate strategies to support learning); and (3) *environmental* (e.g., type of instruction, quality of feedback, nature of interactions with parents and peers). A number of researchers (e.g., Pintrich, 2000; Zimmerman, 2001; Zimmerman, Bandura, & Martinez-Pons, 1992) have demonstrated that students' SRL capabilities can play a significant role in high school academic achievement. In addition, studies by Brown and Palincsar [1989] have demonstrated that through instruction younger students can acquire and apply metacognitive skills, such as planning and monitoring. However, students in typical classrooms are rarely provided opportunities to learn and exercise these strategies (Paris & Paris, 2001; Zimmerman, 1990).

For about 8 years, our research team, the Teachable Agents Group, has been developing computer-based-learning environments that utilize the learning-by-teaching (LBT) approach to instruction in order to foster students' acquisition of knowledge and development of sophisticated metacognitive strategies. The system embodies the social cognitive learning framework and

provides students with opportunities for self-directed, open-ended learning in the domains of science and mathematics (Biswas, Leelawong, Schwartz, Vye, & Vanderbilt, 2005; Blair, Schwartz, Biswas, & Leelawong, 2007; Leelawong & Biswas, 2008). In the system, students are given a *knowledge construction* task in which they engage in the iterative process of reading and building causal concept maps for a range of instructional topics (e.g., climate change, ecology, and thermoregulation). This process is enhanced through the *social interaction* component of the system in which students assume the role and responsibilities of being their agent's teacher. The environment is structured so that successfully instructing their teachable agent ("Betty") requires the students to learn and understand the topic for themselves. Our previous work has shown that students find the task of teaching and interacting with Betty to be motivating, and it also helps them enhance their own learning (Chase, Chin, Oppezzo, & Schwartz, 2009; Schwartz, Blair, Biswas, Leelawong, & Davis, 2007; Schwartz et al., 2009). The teachable agent's performance is a function of how well it has been taught by the student, which provides the student with a non-threatening way of assessing their own understanding and areas of confusion (e.g., "Ugh, Betty is so stupid, now I've got to figure out another way to help her learn this stuff," as opposed to "Why am I not able to get the correct answer?"). Based upon the student's level of progress and pattern of activities, the system triggers responses at appropriate times from Betty or Mr. Davis, the mentor agent, who provides guidance on problem-solving and metacognitive strategies. As a result, the students are

more likely to increase their knowledge of the specific domain content and develop more sophisticated problem-solving and metacognitive strategies, which in turn helps their preparation for future learning (Biswas et al., 2005; Bransford & Schwartz, 1999; Schwartz & Martin, 2004; Schwartz et al., 2007).

This chapter presents analyses from several studies that were conducted in middle school science classrooms, in which students taught their agent about complex science topics, such as river ecosystems and global climate change. One of our goals was to determine the degree to which the agents' metacognitive and SRL prompts could help improve students' learning. Within this framework, we have developed analytical methods to identify and interpret students' learning strategies based on their activity traces in the system. Such analyses can shed light on students' underlying learning processes and the strategies they employ in achieving their learning tasks (Roscoe & Chi, 2007). To date there has been very little work on deriving students' SRL strategies from their activity sequences in computer-based learning environments (some exceptions are Hadwin, Nesbit, Jamieson-Noel, Code, and Winne (2007), Roll, Alevén, McLaren, and Koedinger (2007), and Azevedo, Witherspoon, Chauncey, Burkett, and Fike (2009)). In this chapter, we present a novel methodology that derives HMMs (Li & Biswas, 2002; Rabiner, 1989) from student activity sequences to quantify and assess student learning and metacognition. In addition, we report the results of a second study, where we performed verbal protocol analyses to determine students' acceptance of the strategies discussed by the two agents, and how the feedback provided by the agents influenced their subsequent learning activities.

Measuring Self-Regulated Learning

To effectively design, test, and refine a system promoting SRL skills, it requires the ability to identify and measure metacognitive processes. The traditional approach to measuring students' SRL has been through the use of self-report questionnaires

(e.g., Pintrich, Smith, Garcia, & McKeachie, 1993; Weinstein, Schulte, & Palmer, 1987; Zimmerman & Martinez-Pons, 1986). The underlying assumption in these questionnaires is that self-regulation is an aptitude that students possess. For example, the questionnaire items might attempt to assess students' inclination to elaborate as they read a passage or to determine their approach to managing available time resources (Perry & Winne, 2006; Zimmerman, 2008). This approach has been useful, as the self-report questionnaires have been shown to be good predictors of students' standard achievement test scores and they correlate well with achievement levels (Pintrich, Marx, & Boyle, 1993; Zimmerman & Martinez-Pons, 1986). However, Hadwin and others (Azevedo & Witherspoon, 2009; Hadwin, Winne, Stockley, Nesbit, & Woszczyna, 2001; Hadwin et al., 2007; Perry & Winne, 2006) have argued that while the questionnaires provide valuable information about the learners' self-perceptions, they fail to capture the dynamic and adaptive nature of SRL as students are involved in learning, knowledge-building, and problem-solving tasks.

Increasingly, researchers have begun to utilize trace methodologies in order to examine the complex temporal patterns of SRL (Alevén, McLaren, Roll, & Koedinger, 2006; Azevedo & Witherspoon, 2009; Azevedo et al., 2009; Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010; Hadwin et al., 2007; Jeong & Biswas, 2008; Zimmerman, 2008). Perhaps the most common type of data collected, and the focus of this chapter, is computer logs, which can record every action that the student performs in a computer-based learning environment. An example of computer trace log analysis is presented in Hadwin et al. (2007). They performed a study that collected activity traces of 8 students using the gStudy system (Perry & Winne, 2006). The activity traces were analyzed in four different ways: (1) frequency of studying events, (2) patterns of studying activity, (3) timing and sequencing of events, and (4) content analyses of students' notes and summaries. The results of this analysis were compared against students' self-reports on their SRL. One of the important findings was that many participants' self-reports of studying tactics, as determined by the MSLQ items, were not well calibrated with

studying events traced in the gStudy system. The researchers found that the best matched item showed a 40% agreement, and the average agreement was 27%. The authors concluded from this study that trace data of student activity in e-learning environments are important for furthering our understanding of SRL.

More recently, trace data is being supplemented with other sources of data, such as concurrent verbal think-alouds (e.g., Azevedo & Witherspoon, 2009) and measures of effect (e.g., automatic recording of facial expression and posture) (Burlison, Picard, Perlin, & Lippincott, 2004; D’Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; D’Mello, Picard, & Graesser, 2007; Lester et al., 1997). Azevedo et al. (2009) have developed a hypermedia environment called MetaTutor to help students learn about complex and challenging science topics, such as the circulatory processes in human body systems. The system is also designed to train students in key SRL processes that relate to planning, metacognitive monitoring, learning strategies, and methods for handling task difficulties and demands. The authors used a combination of student trace data and think-aloud protocols to understand the nature of students’ learning outcomes and their deployment of SRL processes. For example, one of their studies showed that students predominantly used strategies that pertained to acquiring knowledge from the multimedia resources, and they only occasionally employ monitoring strategies to check what they have learned (Azevedo & Witherspoon, 2009). Combining trace and think-aloud protocols provides more insight into the students’ thought processes that govern the use of strategies. Furthermore, they can be used to validate the results of the trace data analysis.

Betty’s Brain and Self-Regulated Learning

The Betty’s Brain system, illustrated in Fig. 29.1, implements the LBT paradigm to help middle school students develop cognitive and metacognitive

skills in science and mathematics domains (Biswas et al., 2005; Blair et al., 2007; Leelawong & Biswas, 2008; Schwartz et al., 2007). The system supports five primary types of activities:

- *Read*: The system contains a set of indexed, hypermedia resources that students can access and read at any time while working on the system. These resources contain all of the science information (and more) that students need to build their concept maps.
- *Edit*: Students explicitly teach Betty using a causal concept map representation (Jonassen & Ionas, 2008), where the relevant science concepts are nodes, and causal relations between the concepts are modeled as links. For example, fish eat (decrease) macroinvertebrates and this representation allows students to reason that an increase in fish causes a decrease in macroinvertebrates. Students teach Betty new concepts and links using a visual interface that includes menu selections and templates for adding and modifying information (e.g., the interface contains these four buttons: Teach Concept, Teach Link, Delete, and Edit).
- *Query*: Students use a template, illustrated in Fig. 29.1, to check their teaching by asking Betty questions, which she answers using causal reasoning through chains of links (Forbus, 1984; Leelawong & Biswas, 2008).
- *Explain*: Students can probe Betty’s reasoning, by asking her to explain her answer to a query. She demonstrates the use of causal reasoning processes to derive her answer, and verbalizes her reasoning process using speech and simultaneous animation on the concept map.
- *Quiz*: Students can assess how much Betty has learned by having her take a quiz, which is made up of a set of questions chosen by the Mentor agent. Betty’s inability to answer some of the questions correctly usually motivates the students to learn more so that they can make improvements to the concept map and help Betty do better on her quizzes.

Since our middle school students are novices in the science topics and the teaching tasks, we provide them with a variety of scaffolds to help them overcome obstacles they may face in learning and

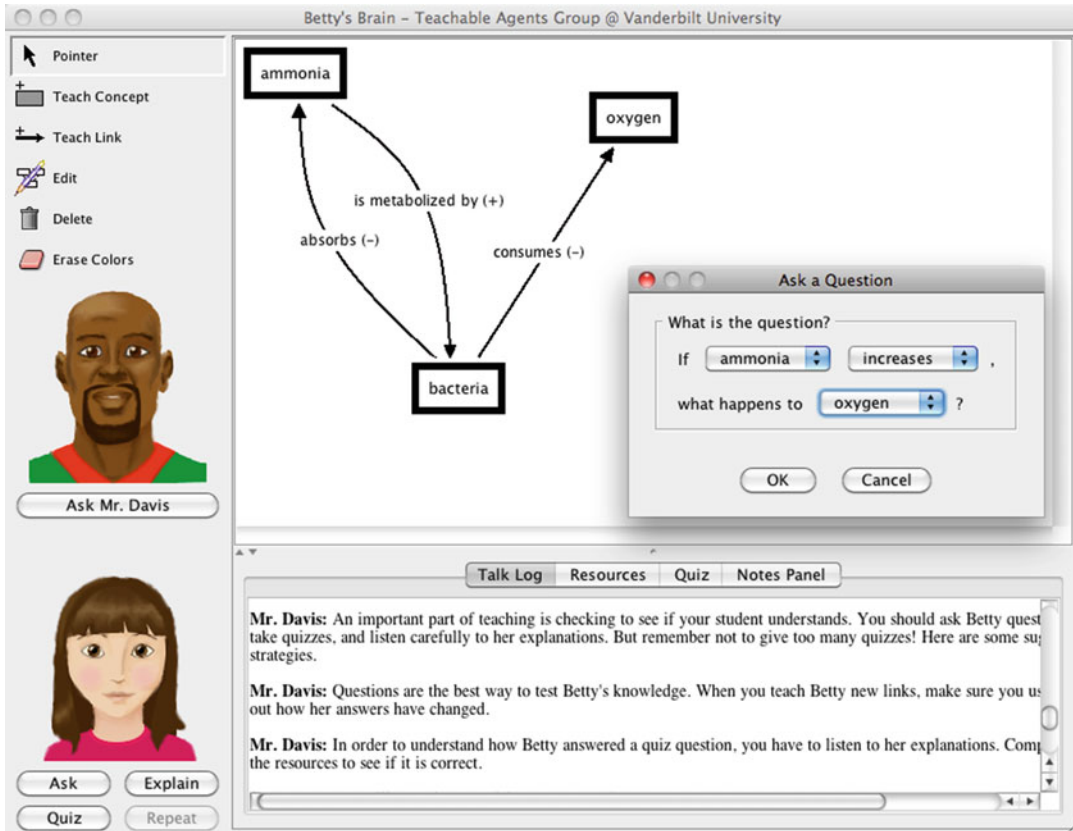


Fig. 29.1 Betty's Brain system with query window

teaching the domain material. In addition to answering queries and taking/administering quizzes, the agents also provide spontaneous feedback to the student on the relative effectiveness of their teaching performance. This feedback is designed to help students develop and employ more metacognitive learning strategies (Schwartz et al., 2007; Tan, Biswas, & Schwartz, 2006; Wagster, Tan, Wu, Biswas, & Schwartz, 2007).

Schunk and Zimmerman (1997) point out that the self-regulation profiles of novice learners are quite distinct from those of experienced learners. Novices are often poor at forethought, and their self-judgment abilities are not well developed. These strategies can be taught, but students in typical classrooms are rarely provided opportunities needed to learn and master them. Our system addresses this problem by adopting a SRL framework that promotes a set of comprehensive skills, such as setting goals for learning new materials

and applying them to map building tasks; deliberating about strategies to enable this learning; monitoring one's learning progress; and revising one's knowledge, beliefs, and strategies as new material and strategies are learned (Azevedo, 2005; Schraw, Kauffman, & Lehman, 2002; Winne & Hadwin, 2008; Zimmerman, 2001).

Figure 29.2 illustrates our conceptual cognitive/metacognitive model that we have employed in designing the Betty's Brain system. Pintrich (2002) differentiates between two major aspects of metacognition for learners: (1) *metacognitive knowledge* that includes knowledge of general strategies and when they apply, as well as awareness of one's own abilities, and (2) *metacognitive control* and self-regulatory processes that learners use to monitor and regulate their cognition and learning. In our model, metacognitive control is illustrated in the monitoring and knowledge construction strategies in Fig. 29.2. In more detail,

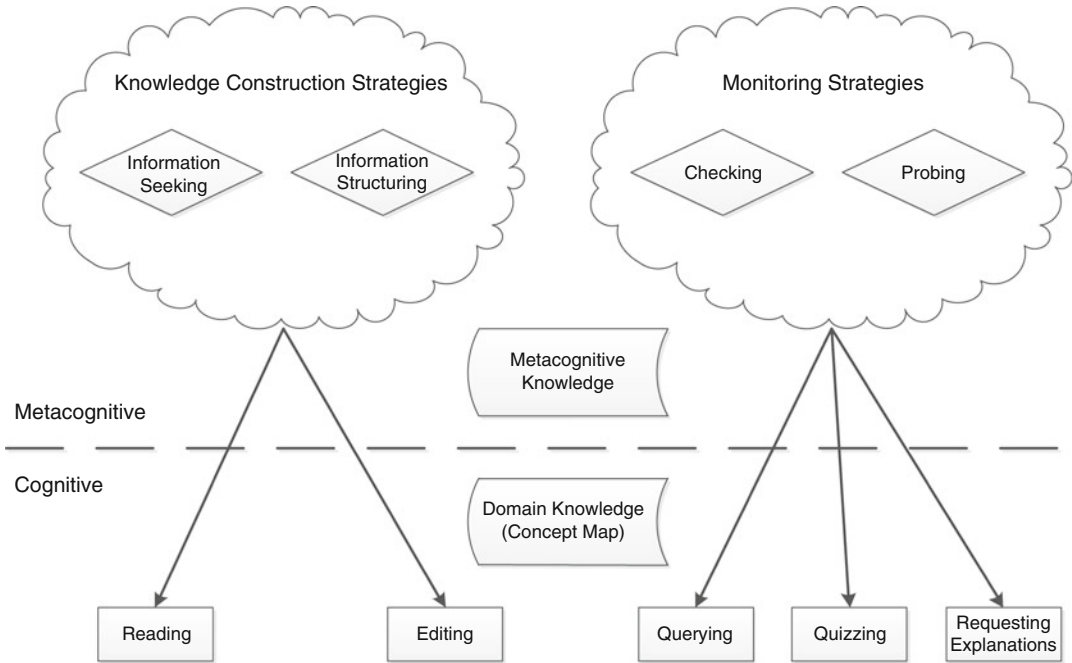


Fig. 29.2 Our model of self-regulated learning strategies and activities in the Betty's brain system linked to these strategies

Pintrich discusses a goal orientation framework for characterizing SRL that covers mastery and performance orientations to achieving goals (Pintrich, 2000). In our approach, feedback from the mentor promotes mastery orientation, e.g., focus on learning with understanding, and setting standards for checking and probing the map (asking queries and reflecting on the explanations generated by Betty) to make sure it has no errors. Betty's interactions with the student focus more on the avoidance aspect of mastery orientation, i.e., making sure students strive for self-improvement and work toward producing an error-free map.

For knowledge construction in the Betty's Brain system (i.e., building causal concept maps), we identify two key types of mastery-oriented self-regulation strategies: (1) *information seeking*, in which students study and search available resources in order to gain missing domain information or remediate existing knowledge, and (2) *information structuring*, in which students structure the information gained into causal and taxonomic relationships to build and revise their concept maps. Information seeking strategies are

directed toward effective use of the resources in the system, whereas information structuring focuses on strategies for construction and revision of the concept map.

The model also posits two types of monitoring strategies: (1) *checking*, where students use the query or the quiz features to test the correctness of their concept map and (2) *probing*, a stronger monitoring strategy, where students systematically analyze their map in greater detail, by asking for explanations and following the causal reasoning steps generated by the agent to locate potential errors. Effective guidance (i.e., relevant and timely feedback) based on this SRL model makes students aware of their learning strategies and helps them develop better strategies, such as rereading the resources to check if there are errors in their concept maps (combining information seeking and checking strategies), and asking queries and checking explanations to find the source of an error (a probing strategy).

Table 29.1 provides examples of the agent feedback, which is triggered by students' activity patterns (see column 2) and linked to strategies for

Table 29.1 Examples of agent responses to observed student behavior patterns

SRL strategy	Triggering activities	Betty response	Mr. Davis response
Knowledge construction: information seeking	Student taught Betty several concepts, but has added only a few links.	<i>What do the resources say about that concept? Could we search the resources to learn more?</i>	<i>Reading is a very important part of learning. You should read the resources often to help you understand and double-check what you are teaching. Here are some strategies that you can use to improve your reading skills.</i>
Knowledge construction: information structuring	Student has just added a number of concepts but no links to the concept map.	<i>Can we read the resources to make sure we understand the relations between these concepts?</i>	<i>A good teacher explains to her students how concepts affect each other. Add links to the map to teach Betty how the concepts cause other concepts to change.</i>
Monitoring: checking	Betty has incorrect answers on the quiz she just took.	<i>Could you ask me some review questions to see how good my answers are?</i>	<i>Questions are the best way to test Betty's knowledge. When you teach Betty new links, make sure you ask her questions to find out how her answers have changed.</i>
Monitoring: probing	Betty has got some of her quiz answers wrong.	<i>These quizzes can be tough. Can we go over my explanations to see how well I understand?</i>	<i>After each quiz, try to review a few of Betty's answers and explanations. Follow each step in her explanations carefully to be sure that they make sense.</i>
Monitoring: pointing out a suboptimal strategy	Student asks Betty to take repeated quizzes without reading or asking questions.	<i>I want to do well on the quizzes. It would help if I learned more between each quiz instead of taking so many.</i>	<i>You have been asking Betty to take a lot of quizzes recently. Try asking her questions to see where she is making mistakes, and teach her more in between quizzes.</i>

knowledge construction and monitoring implied by our model.¹ The agents have different roles (and relationships with the student) in the system, which affects the wording and the content of the feedback they provide. Betty's persona and role as an engaged student "interested in learning and performing well," is influenced by the social cognitive framework. Betty's feedback incorporates metacognitive awareness that she conveys to the students at appropriate times to help them develop and apply monitoring and self-regulation strategies (Schwartz et al., 2009; Wagster et al., 2007). Mr. Davis, the mentor, and, therefore, the more knowledgeable persona in the system, provides help in the form of suggested activities linked to

effective SRL strategies (e.g., "if you are not sure, check the resources to see if Betty is answering her questions correctly.").

Experimental Studies

We have conducted several classroom studies where students use the teachable agents system to learn and gain a better understanding of a variety of science topics, such as river ecosystems, thermoregulation, and climate change. In these studies, the topics and specific science content provided by the system are closely linked to the middle school science curriculum. At the beginning of each study, the science teacher introduces students to the topic during regular classroom instruction. The intervention phase starts with an overview of causal relations and causal mapping during a 45-min class period. This is followed by a hands-on training session with the system the next day. Over the next 4 or 5 days, the students teach Betty by building a causal concept map for the science topic, which represents what Betty knows.

¹In the system, the same triggering conditions are used to generate Betty's and Mr. Davis' feedback. The system is designed so that the feedback is provided only after the triggering pattern is activated a certain number of times. This number is chosen randomly from a predefined range of values (e.g., [2,5]) and recomputed after every instance of feedback. The numbers for Betty and Mr. Davis are chosen independently.

To assess students' acquisition of science domain knowledge and causal reasoning skills, we employ two measures. The first is a pretest to posttest gain score. These tests contain a mix of content-related multiple choice and free response items (Biswas et al., 2010; Leelawong & Biswas, 2008) that are administered before the students are introduced to causal reasoning, and at the end of the intervention. The second measure examines students' final maps, in terms of completeness and accuracy.

In this chapter, we analyze the results from two classroom studies. The first study compared the students' use of SRL strategies in three different conditions described below. We had two questions: (1) Would students who taught an agent use more SRL strategies in their learning and teaching tasks than students learning entirely for themselves? and (2) Would students who received SRL feedback from the agents use more sophisticated SRL strategies than students who did not? The second study used verbal protocol analysis to assess the effectiveness of different kinds of SRL strategies, and also checked whether the feedback provided by one agent was more effective than the feedback provided by the other agent. The results, and a discussion of these results, are presented in the remainder of this section.

Study 1: Modeling Students' SRL Strategies

In this study, our goal was to determine if teaching the Betty agent and providing metacognitive feedback would help students become better learners than those who did not teach or receive the feedback. Our participants were 56 students in 2 fifth-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 river ecosystem concept maps over five 45-min sessions. Two of the conditions (1) the LBT group and (2) the self-regulated learning-by-teaching (SRL) group created their map to teach Betty so that she could pass a test on her own. In addition to the teachable agent, both groups had access to Mr. Davis, the mentor agent.

As students taught Betty, they could ask her questions, get her to explain her answers to the questions, and take quizzes, which were sets of questions created by Mr. Davis. After Betty took a quiz, the mentor graded the quiz and displayed the results to the students. Both systems also provided feedback to students after a quiz.

The differences between the LBT and SRL groups were in the feedback provided. In the LBT version of the system, Mr. Davis provided *corrective* feedback after the quiz results were displayed. The corrective feedback was linked to a quiz question that produced an incorrect answer, and it included information about one of the following: (1) a missing concept that would be required to generate the correct answer; (2) a missing link that would be required to generate the correct answer, or (3) a link that was incorrectly represented in the map (e.g., one of the link effects was incorrect, or the direction of a link was reversed). The mentor's feedback would first pick on missing concepts, then missing links (i.e., if the student's map contained the relevant concepts to answer the question), and last, incorrect links (i.e., if all necessary concepts and links were on the map, but one or more links were incorrectly specified or extraneous).

In contrast, the SRL version of the system provided the SRL strategy feedback presented in Sect. 29. After seeing Betty's quiz results, the students could ask the mentor for suggestions. In response, Mr. Davis would suggest relevant SRL strategies, such as an information seeking strategy: he would point to keywords for finding relevant sections of the resources to learn more about concepts and relations that were missing/incorrect in the map. In addition to feedback after a quiz, Betty and Mr. Davis also generated spontaneous responses triggered by the activity patterns, such as the ones described in Table 29.1.

Our control condition for the study, the intelligent coaching system (ICS) group was told to create the map to learn for themselves. The Betty agent was removed from this version of the system, and the students interacted only with the mentor, Mr. Davis. Otherwise, the activities available in the ICS interface were identical to the two LBT systems. For example, students in the ICS group could also query their map and ask for

explanations, but in this case, it was Mr. Davis, and not Betty, who responded to them. Similarly, ICS students took the quiz for themselves rather than having Betty take the quiz. The content and form of quizzes and explanations were identical for the ICS, LBT, and SRL groups. In the ICS group, Mr. Davis provided the same corrective feedback as in the LBT version of the system.

All student activities in the system were captured in log files. Each activity was assigned to one of five primary categories: (1) EDIT—add, edit, or delete concepts and links in the concept map; (2) QUER(y)—query Betty on a portion of the map; (3) QUIZ—ask Betty to take a quiz; (4) READ—read the resources; and (5) EXPL(anation)—ask Betty to explain her answer to a query. For each activity, the program captured additional information related to the activity. For example, when the student asked a question, the question and Betty’s response to the question were also stored in the log file.

Analyses that do not take into account the sequential nature of student interactions with the system, such as counting the frequency of student activities, can provide only limited information for learning strategy models of student behavior (Biswas et al., 2010). We believe that a state-based representation that captures the sequential characteristics of students’ activities provides a more powerful narrative of the student learning behaviors. HMMs (Rabiner, 1989), which contain a set of states and probabilistic transitions between those states (more likely transitions are assigned higher probabilities), provide such a representational scheme. The states in a HMM are hidden, meaning that they cannot be directly observed in the environment/system. Instead, they produce output (e.g., student activities in the Betty’s Brain system) that can be observed. Deriving a HMM from activity traces requires simultaneous estimation of (1) the number of states; (2) the probabilities associated with transitions between states; (3) the probabilities associated with observing certain outputs (i.e., particular student activities, such as reading or querying activities); and (4) the probability of a state being the initial state in an activity sequence.

By providing a concise representation of student learning strategies and behaviors, HMMs

have the potential for providing a high-level view of how students approach their learning tasks (e.g., what strategies they use and how they switch between strategies) (Biswas et al., 2010; Jeong & Biswas, 2008). Algorithms for learning an HMM from output sequences are well known but require appropriate configuration/initialization parameters for effective use (Rabiner, 1989). Specifically, HMM learning algorithms require an initial HMM description, whose parameters are then modified to maximize the likelihood of producing observed output sequences. In particular, the number of states in the HMM and their initial output probabilities can have a significant effect on the resulting, learned HMM.

We have developed an algorithm designed to generate HMMs from a set of student activity sequences (Jeong & Biswas, 2008; Li & Biswas, 2000, 2002). The first step in the analysis is to extract each students’ activity sequences over the period of the study from the log files. Although all students had access to the full set of actions, not all of them used them effectively. Using queries to check whether recent revisions to the map were correct, or to locate errors in the concept map, is an example of effective use of queries. On the other hand, asking questions simply to make Betty speak, so that the student could make fun of her mechanical, computer-generated voice is clearly an ineffective use of queries for the learning task. When students generated questions that were not related to parts of the map they had worked on recently, it was unclear whether these queries were related to effective learning. We addressed this issue by developing a *relevance score* that took into account how much the current action could be linked to other recent actions.

Each student action was assigned a relevance score that depended on the number of relevant previous actions within a pre-specified window. This score provides a measure of *informedness* for knowledge construction activities and, similarly, a measure of *diagnosticity* for monitoring activities. Overall, the relevance score provides a rough measure of strategy consistency or coherence over a sequence of actions. For this analysis, a prior action was considered relevant to the current action if it was related to, or operated on, one

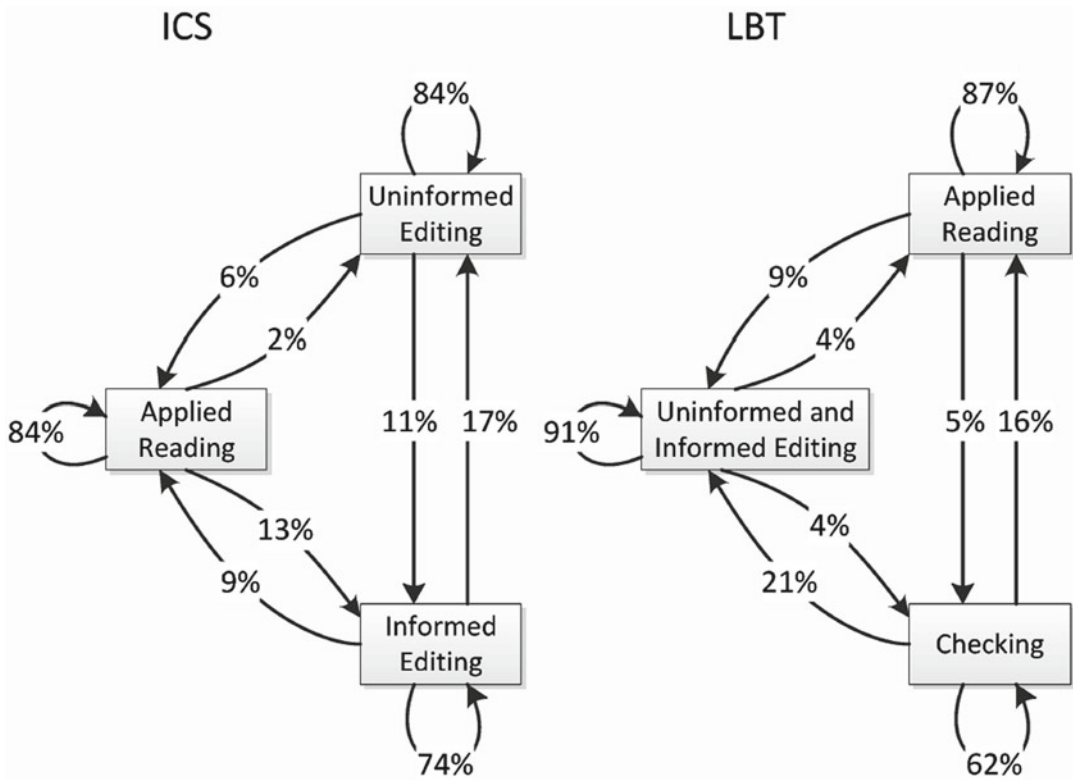


Fig. 29.3 ICS and LBT group HMMs derived from activity sequences

of the same map concepts or links. For example, if a student edited a link that was used to generate an answer in a recent query, the query action was counted in the edit’s relevance score. The increased relevance score suggested a more informed edit action because it was related to a recent query.

The relevance score is employed in HMM generation by refining the classification of student activities. Each of the actions in an activity sequence is assigned a label, H (high) or L (low), based on its relevance score, in order to maintain the context and relevance information of the actions in the sequence. For example, a QUER-H activity implies that the query the student asked is related to other activities recently performed, while a QUER-L implies that the query activity is largely unrelated to the students’ recent activities.

The HMM models derived for the ICS, LBT, and SRL groups are shown in Figs. 29.3 and 29.4. States in the models are named based on an interpretation of their outputs (activities) illustrated in

Figs. 29.5 and 29.6. The possible transitions between states are shown as arrows, and the transition probabilities are expressed as percentages. For example, the ICS behavior model indicates an 84% likelihood that a student who just performed an applied reading action (i.e., one of the observable actions associated with the Applied Reading state described below) will next perform another applied reading action, but there is a 13% chance that the student will perform an informed editing action (i.e., an action produced by the Informed Editing state) next. The models for the ICS and LBT groups each have three states, but the activities associated with some of those states differ significantly. Therefore, the states are interpreted, and named, differently for those groups. Further, the derived model for the SRL group has five states instead of three and shows some interesting differences in the set of actions associated with those states.

We used the activities associated with a state to categorize the states of the three derived HMM

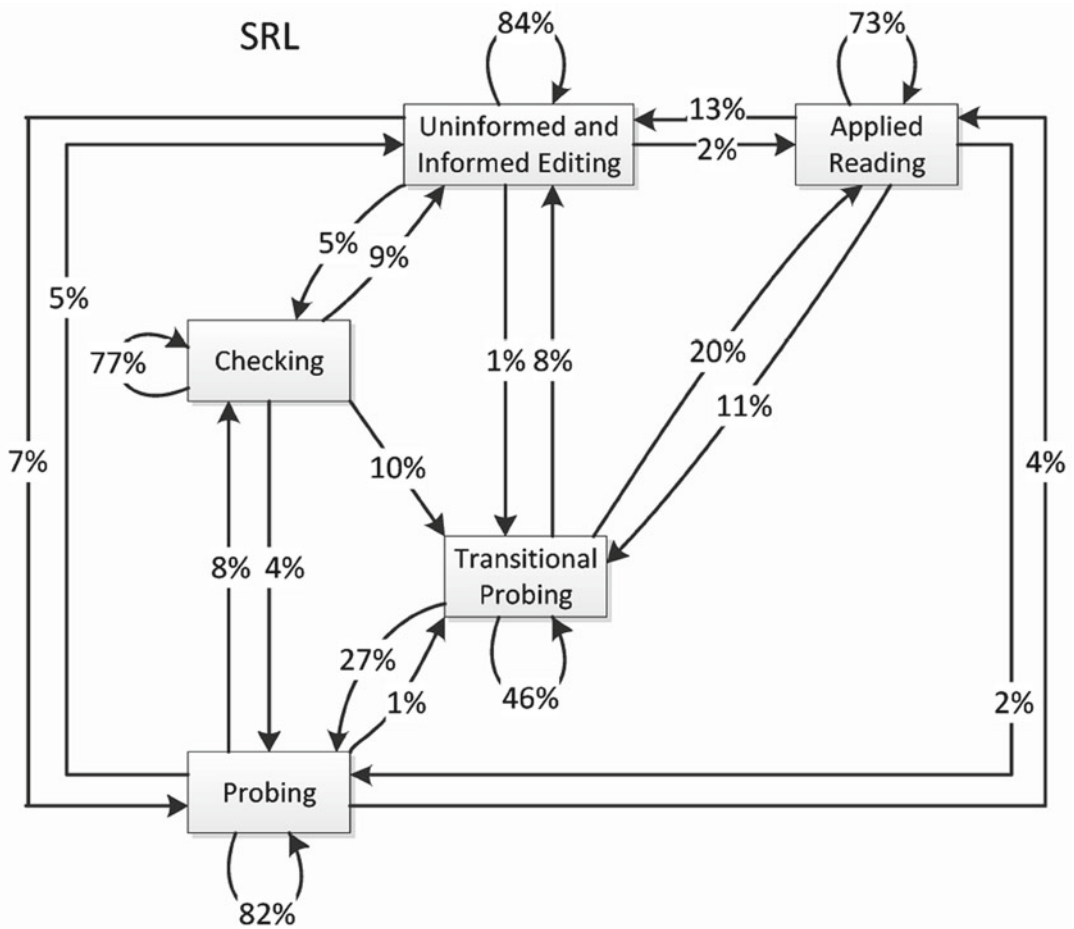


Fig. 29.4 SRL group HMM derived from activity sequences

models. This analysis produced seven different types of states that are described below.

1. *Applied reading*—students are primarily engaged in reading the resources and applying the knowledge gained from reading by editing their maps. This state combines information-seeking strategies with informed information structuring.
2. *Uninformed editing*—students are primarily making uninformed changes to their map, indicating the use of trial-and-error or guessing strategies for information structuring. Students may generate queries, but the queries generally do not relate directly to the editing activities. This represents a suboptimal information structuring strategy.

3. *Informed editing*—students are primarily making informed changes to their map (information structuring) based on relevant queries or quiz questions. As opposed to uninformed editing, the students are using queries and quizzes to guide their map editing actions.
4. *Uninformed and informed editing*—students are primarily making changes to their map, some of which are based on relevant queries or quizzes. This state combines the activities of the uninformed editing and informed editing states, including situations where students are making edits relevant to recent queries and quizzes, as well as situations in which students are making edits without focusing on a single area of the map.

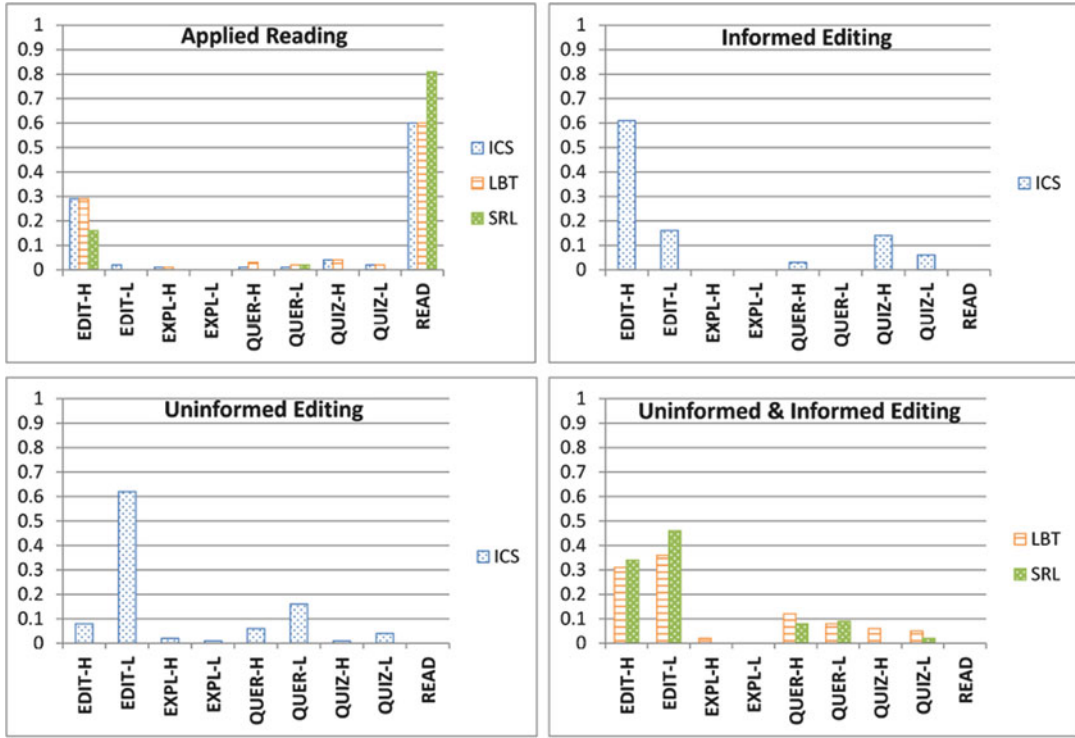


Fig. 29.5 Activities in knowledge construction states

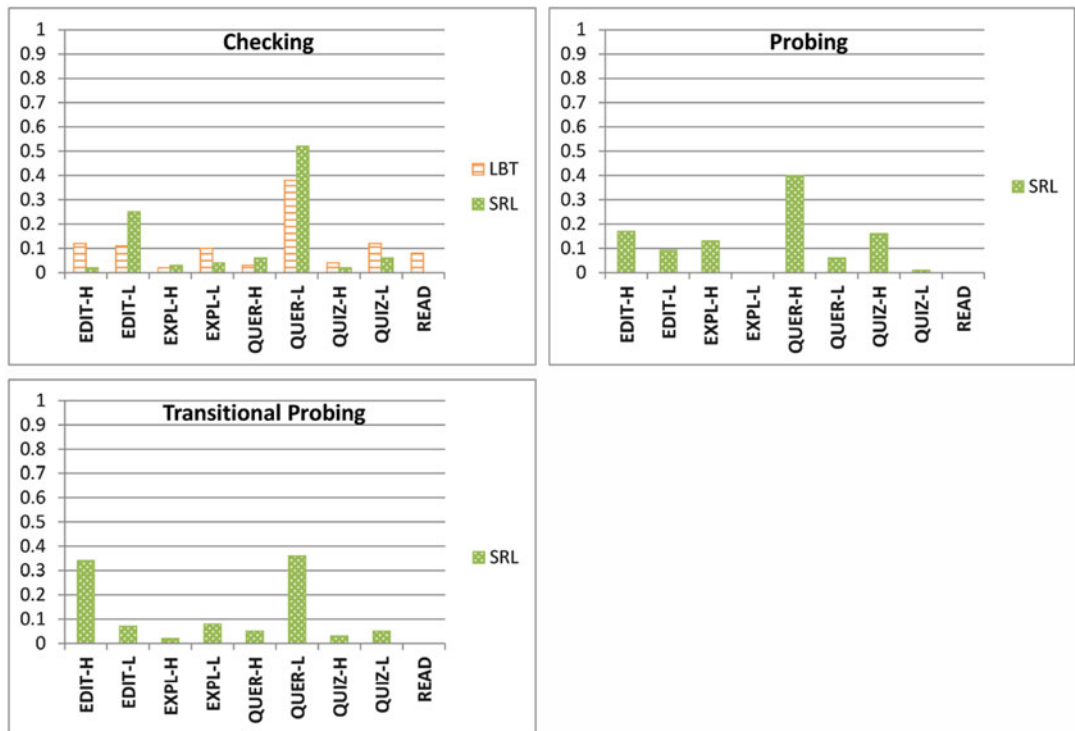


Fig. 29.6 Activities in monitoring states

Table 29.2 Proportion of expected state occurrences by condition

Behaviors	ICS	LBT	SRL
	Proportion (%)	Proportion (%)	Proportion (%)
Applied reading	33	30	17
Uninformed editing	36	–	–
Uninformed and informed editing	–	60	34
Informed editing	31	–	–
Checking	–	10	13
Transitional probing	–	–	7
Probing	–	–	29

5. *Checking*—students are querying and quizzing Betty to check the correctness of their concept maps. However, the use of queries and quizzes may be unfocused. For example, queries may not be related to recently edited areas of the map, and it is not clear that students are using the quiz results to focus on areas on the map where there are errors. Therefore, this state corresponds to a weak monitoring strategy.
6. *Probing*—students combine querying and quizzing with the explanation feature, which illustrates the chain of links that were followed to generate an answer to a question. Further, the queries, explanations, and quizzes are focused on a particular area of the map, and the results inform map editing. This combination implies a deeper, more focused monitoring strategy than the checking state and may be evidence of metacognitive reflection on the quality of the student’s map/knowledge.
7. *Transitional probing*—students perform activities similar to the probing state, but generally with lower relevance scores, suggesting that they may be transitioning to probing a different area of the concept map.

As discussed above, each of the interpreted states can be mapped onto one or more knowledge construction and monitoring strategies outlined in our conceptual SRL model that was illustrated in Fig. 29.2. The HMMs provide evidence that the SRL condition uses more effective monitoring strategies (i.e., probing strategies in addition to checking strategies) than the LBT and ICS conditions.

We probed further to determine the prevalence of individual states suggested by a generated HMM. To do this, we calculated the proportion

of *expected state occurrences* by condition in Table 29.2. This calculation uses the HMM to provide an *expected value* for the average frequency with which a state would occur when producing sequences of a given length. Specifically, the expected state occurrences measure employs state transition probabilities in the derived HMM and average activity sequence lengths from the trace data to calculate an expected value for the proportion of individual state occurrences (Biswas et al., 2010). Although states corresponding to knowledge construction behaviors account for a significant percentage of behaviors in all groups, the HMMs for the LBT and SRL groups also show use of monitoring strategies (10% for LBT and 49% for SRL). The SRL HMM also includes more states suggesting a greater number (and possibly greater complexity) in the types of strategies employed. Further, the activities involved in these additional states suggest use of *probing*, a more advanced monitoring behavior, which is absent from the ICS and LBT HMMs.

The results of the HMM analysis identify differences in strategies employed by the different groups of students, but do not directly indicate the effect of these behaviors on student learning. Therefore, Table 29.3² shows the learning gains measured by tests and map scores for each

²All statistical comparisons of means among conditions were made with ANOVA post-hoc (Tukey HSD) tests, and effect sizes are computed as Cohen’s *d*. Further, since some of the differences falling outside of the significance cutoff of $p < 0.05$ still had moderately large effect sizes, we report the results for multiple significance cutoff values ($p < 0.1$ and $p < 0.05$), allowing the reader to make their own determinations based on the reported results.

Table 29.3 Mean pre-to-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)

condition in the study. Results indicate that the two groups that taught Betty (LBT and SRL) outperformed the ICS group on gains in both test and map scores, although not all of these differences were statistically significant. In particular, differences on multiple choice test score gain were not statistically significant between any of the conditions. However, for the free response test questions, the SRL group showed greater gains than the ICS group ($p < 0.1$ and a moderately large effect size of $\hat{d} = 0.72$). For the gain in correct map concepts, the SRL group outperformed both the ICS group ($p < 0.05$, $\hat{d} = 0.81$) and the LBT group ($p < 0.01$, $\hat{d} = 1.05$). Similarly, for the gain in correct map links, the SRL group outperformed the ICS and LBT groups ($p < 0.05$, $\hat{d} = 0.97$ and $p < 0.1$, $\hat{d} = 0.72$, respectively).

Overall, these results indicate that the students who taught Betty (i.e., LBT and SRL groups) outperformed the other students (i.e., the ICS group), both in learning gains and the use of monitoring strategies. Although the SRL group received different feedback (SRL rather than corrective) from the mentor, the only difference between the LBT and ICS groups was whether students taught Betty or learned for themselves. The ICS students use of less effective learning strategies, as apparent in the HMMs, may explain their smaller learning gains. Further, the SRL group had higher free response and map score gains than the LBT group (although not all of the differences were statistically significant for the number of students in this study), suggesting that the SRL feedback promoted more effective learning and concept mapping performance. Moreover, while 60% of SRL students completed their concept maps during the five sessions, only 44% of LBT students and 31% of ICS students were able to complete

their concept maps. The results of the HMM analysis, combined with the results on learning gains, suggest that the metacognitive feedback helped students implement SRL strategies, which allowed them to more effectively learn the science content. Although the HMM analysis illustrates the effectiveness of providing metacognitive feedback in Betty's Brain, it does not indicate which agent or types of feedback were most effective. This was the focus of the second study, which we describe next.

Study 2: Comparing the Mentor and Teachable Agent Feedback

In order to assess the effectiveness of different forms of feedback in our system (i.e., differences by (1) agent and (2) content of feedback: knowledge construction versus monitoring), we conducted a study, which included a think-aloud protocol to determine students' reactions to the agent feedback. The study was conducted in 3 fifth-grade science classrooms in the same school as study 1.³ Two of the classrooms had the same science teacher as in study 1. The third classroom had a different teacher, but teacher 2 worked closely with teacher 1 for the unit taught in this study. All students worked on a newer version of the SRL system from study 1. In this version of the system, the feedback from the two agents was better organized into the categories described in Table 29.1.

Students worked in a total of 40 pairs chosen by the teachers to ensure that the paired students were at similar academic levels and had compatible personalities. Before the study began, the teachers instructed students on how to collaborate on the system. The students had to discuss with one another and come to a consensus before they performed an action on the system. Control of the keyboard and mouse alternated between the students (e.g., if one student had control of the input devices on day 1, then the partner was given control on day 2). The science teachers

³Study 1 and study 2 were conducted in different years.

Table 29.4 Student verbal response to agent feedback

Agent →	<i>N</i>	Referenced (%)	Affirmed (%)	Dismissed (%)	Deferred (%)
Betty	649	34	6	16	2
Mr. Davis	275	30	18	10	2

ran a brief practice session on working with the system before the students started this phase of the study. All students had worked individually with the Teachable Agent system on another science unit (ecosystems), so they were familiar with the system.

Students worked on the topic of pollution in river ecosystems for three 45-min periods. We recorded student conversations and interactions using webcams. After the study was concluded, two coders reviewed all of the video data and recorded students' responses to the feedback. For every instance in which the TA or the mentor provided feedback, the coders noted whether the students' subsequent discussion *affirmed*, *dismissed*, or *deferred* the agent's feedback. Inter-rater reliability for each category was over 85% with Cohen's kappa values over 0.6, and the results are summarized in Table 29.4.

Students explicitly referenced the feedback from the agents about a third of the time (34% for Betty and 30% for Mr. Davis). Even when students did not explicitly reference the feedback, they sometimes responded to the feedback by talking directly to the agent in response or suggesting a course of action directly indicated (or contra-indicated) by the feedback. All student discussions following feedback were coded in three categories of possible response to the feedback⁴: (1) affirm (e.g., "We should do that" or "We need to read more" responding to feedback suggesting the students read the resources), (2) dismiss (e.g., "No, I don't want to read" or "Let's just keep giving her quizzes" responding to feedback suggesting students teach Betty more between giving her quizzes), and (3) defer (e.g., "Hold on, we will get to that in a second").

As illustrated in Table 29.4, there were differences between how frequently the students affirmed or dismissed feedback from the two agents. Students were more likely to affirm feedback from Mr. Davis, and were more likely to dismiss feedback from Betty. This suggests that students paid less attention to the self-reflective feedback from Betty than to the more explicit, strategy-oriented feedback from Mr. Davis. Although one possible explanation for this difference is that Mr. Davis provided better feedback and advice, the tenor of student discussions indicated that they treated Betty like a less-knowledgeable peer, while according Mr. Davis the status of a knowledgeable authority figure and considering his advice more carefully.

To understand how students' verbal responses related to learning, we analyzed the study results for the two metacognitive categories of feedback from each agent: (1) knowledge construction strategies, and (2) monitoring strategies. Table 29.5 summarizes the percentages of each type of verbal reaction to the different forms of feedback, as well as their correlation with the student pair's final map score. Students who more frequently affirmed the knowledge construction strategy feedback from either the TA or the mentor had higher map scores, but the correlations were not statistically significant. Students who dismissed either the knowledge construction or the monitoring feedback from either agent had lower map scores (negative correlations). However, when the students affirmed the monitoring feedback, the results were surprising. Affirming Mr. Davis's monitoring feedback showed a positive correlation with map score (not statistically significant), but affirming Betty's monitoring feedback was negatively correlated with map score ($p < 0.05$). We discuss this result in greater detail later, but overall the students seemed to affirm the knowledge construction feedback more, and affirming this feedback implied higher map scores.

⁴Many student discussions, including some that explicitly referenced the feedback, neither affirmed, dismissed, nor deferred the feedback.

Table 29.5 Verbal responses to feedback and corresponding map score correlations (^b $p < 0.05$)

Agent → Category ↓	Betty			Mentor		
	Percent verbal response			Percent verbal response		
	(Correlation with map score)			(Correlation with map score)		
	Affirm	Dismiss	Defer	Affirm	Dismiss	Defer
Knowledge construction	9% (0.20)	27% (-0.45 ^b)	2% (-0.16)	20% (0.37)	22% (-0.31)	2% (0.24)
Monitoring	4% (-0.46 ^b)	11% (-0.35)	2% (0.04)	17% (0.28)	6% (-0.44 ^b)	2% (-0.16)

The verbal responses to feedback listed in Table 29.4 suggest a difference in the way students react to the feedback from the two agents, and these reactions affect their concept map building performance.⁵ For example, those who affirmed Mr. Davis’s knowledge construction and monitoring feedback seemed to do better in their map building task.

To determine whether the students’ verbal responses to feedback matched their expected actions in the system, we analyzed student actions immediately following each agent feedback statement. For example, if Betty said “Can we go over my explanations to see if I am missing anything,” we checked to see if the subsequent student actions included asking Betty to explain an answer. Table 29.6 reports for both Betty and Mr. Davis (1) the average number of feedback events by category per student, (2) the average proportion,⁶ of subsequent activities that matched the actions advised by the feedback (using a window size of 3 actions)⁷ and (3) the correlation between the percentage of matched actions and the students’ final map scores.

Overall, the correlation between percentage of matching actions (out of the three student actions subsequent to the feedback) and the students’ final map scores was positive (0.34 for Betty and 0.36 for the Mentor), but the correlations were not statistically significant. More detailed analysis by category of metacognitive feedback, showed a positive correlation between students’ final map scores and their following Betty’s and Mr. Davis’s advice on knowledge construction feedback. Students were more likely to follow Betty’s feedback suggestions than those of the Mentor, but the differences were small (28–24%). The more students’ subsequent actions matched the feedback, the higher their map scores, as measured by the correlations: 0.52 for matching Betty’s knowledge construction advice ($p < 0.1$) and 0.11 for matching the Mentor’s knowledge construction advice. These results differ from the verbal responses to feedback, where students affirmed Mr. Davis’s knowledge construction feedback more than they did Betty’s, and the corresponding correlations with map scores were also higher for the mentor (0.37 versus 0.2).

On the other hand, for monitoring feedback, Mr. Davis appears to have been more effective than Betty. Though the relative number of Betty monitoring feedback events was high compared to Mr. Davis’s (13.1–5.5), student actions after Betty’s feedback showed a poor match to the feedback content (only 0.5%). For the mentor feedback the match was 33%. Combining this information with the verbal response results indicates that the students were more dismissive of Betty’s monitoring feedback, and at the same time they rarely followed up with activities that matched the feedback content. In addition, the correlation between the activity match percentage and students’ map scores was negative, implying

⁵Agent role (and consequently relationship with the student) and the content of agent feedback are inextricably linked in this study, making it impossible to attribute student responses to one factor or the other. However, the correlation between students’ responses to agent feedback has useful implications for future system design and experimental study opportunities discussed in this section and the next.

⁶We employed a proportion in a window of subsequent actions because agent feedback often suggested a course of action that could involve repeated actions (e.g., edits or reads), and it is not possible to determine precisely whether a student’s action was an attempt to follow agent advice or not.

⁷We tested a variety of different window sizes, and all of them produced similar results.

Table 29.6 Action response to feedback and corresponding map score correlations ($^*p < 0.1$)

Feedback category → Agent ↓	Measures	Knowledge construction	Monitoring
Betty	Feedback events	5.77	13.08
	Action (proportional) Match	28.2%	0.5%
	Map score correlation	0.52*	-0.41
Mr. Davis	Feedback events	2.46	5.54
	Action (proportional) Match	24.02%	33.0%
	Map score correlation	0.11	0.26

those who affirmed Betty's monitoring feedback or tried to apply it ended up with lower map scores. On the other hand, Mr. Davis's monitoring feedback had more affirmations and there were more attempts to follow his suggestion, and these correlated with higher map scores (though the correlations were not statistically significant).

Together the verbal and action response results show clear differences in the way students responded to the two agents. Overall, the students affirmed the mentor's feedback more than they did the teachable agent's, and in general, higher affirmation levels implied better final map scores. These results also indicate that the monitoring feedback was less effective than knowledge construction feedback. With the exception of Betty's monitoring feedback, the results showed positive correlations with map scores for both verbal and action response measures. Although there are many potential explanations for the negative correlation between Betty's monitoring feedback and map scores, the results suggest that her monitoring feedback was generally ineffective in helping students improve their concept maps. This could imply that students could not understand Betty's feedback and, therefore, they did not apply her suggestions during their learning and teaching tasks. The few who did, may have applied them inappropriately, and, therefore, used up time that could have been more productively spent in other activities. Alternatively, this could have been the result of self-selection, in which lower-performing students attempted to apply Betty's monitoring advice even though they did not understand it. However, those who followed similar feedback from Mr. Davis did better on their map. Overall, the results indicate that the metacognitive feedback had a generally positive effect on students' learning, but the more

explicit strategy feedback from the Mentor agent was more effective than Betty's self-evaluative statements and suggestions.

Discussion and Conclusions

The Betty's Brain system is designed to leverage the benefits of learning by teaching and causal reasoning to help students learn science. The teaching interactions and agent feedback support students' engagement and promote the development and use of educationally productive cognitive and metacognitive processes. In study 1, students who utilized learning by teaching versions of our system (i.e., the LBT and SRL groups) constructed better concept maps than students who used the non-teaching ICS version of the system. Moreover, students' performances were strongest when the system explicitly supported their use of SRL strategies by having Betty model and prompt for such behaviors, and having the mentor provide additional strategy-oriented advice.

Our approach to analyzing students' activity sequences using HMMs produced good results. We were able to characterize students' activity patterns into a number of (good and bad) knowledge construction and monitoring strategies. The interpretation of SRL group behavior with the HMMs also matched the SRL feedback model we implemented in the Betty's Brain system, while the LBT group HMM showed only one of the two types of monitoring strategies (i.e., checking behaviors) and the ICS group HMM did not show either of the monitoring strategies.

Although the HMM analysis illustrated the effectiveness of providing metacognitive feedback in the Betty's Brain system, it did not indicate which agent or types of feedback were most

effective in promoting SRL behaviors. Our second study included a think-aloud protocol to determine students' reactions to the agent feedback. We combined the think-aloud protocols with analysis of student activity traces to develop a more complete picture of how well students employed the feedback to their map building tasks. Overall, students' verbal responses to agent feedback suggested that they were more receptive to the explicit, strategy-oriented advice from the mentor agent, as opposed to the self-reflective, but less explicit, feedback from the teachable agent. Further, students were more likely to affirm the knowledge construction feedback from each agent than the monitoring feedback. This analysis also showed a positive correlation between affirming feedback and students' map scores, except in the case of Betty's monitoring feedback.

Additional analysis of student responses to feedback, in terms of actions taken following feedback events, showed a similar differentiation between knowledge construction and monitoring feedback. Students taking more actions consistent with an attempt to apply the knowledge construction feedback tended to have better map scores. However, students taking more actions advised by Betty's monitoring feedback tended to have lower map scores, suggesting they were unable to apply the strategies suggested by that feedback. This brings up a number of issues. It suggests that students find it easier to understand and apply knowledge construction strategies (e.g., read the resources to find the correct relation between two concepts or check the resources to see if all of the required concepts appear on the map) than monitoring strategies (e.g., ask a query to check if the map is correct or ask for an explanation to check an answer step-by-step to identify errors). Other studies, such as Azevedo et al. (2009) also suggest that students rarely employ monitoring strategies during learning and knowledge construction tasks, but frequently apply a variety of other metacognitive strategies.

It may also be true that students understand a monitoring strategy but do not know when to apply it, since the feedback only implicitly addressed this issue by advising strategies at appropriate times. For example, when constructing

their concept map, students may not know when to switch from map building to map checking and back in an effective way. Moreover, they may have difficulty in formulating "good" queries that help them check a relevant part of their map. Therefore, monitoring strategy feedback may need to be presented in more elaborate detail with justification of its importance in the learning task and identification of applicable situations. For example, analysis of the context and details of advised actions (e.g., Betty's feedback "Can we go over my explanation step by step and check it with the resources?") suggests the use of explain-and-read actions, but effective application of the feedback involves reading sections of the resources related to the map concepts and links in the current query and explanation. Some of these details may need to be built into the feedback mechanisms, especially in the early stages, to help students learn *when* and *how* to apply strategies in an effective way.

Since students appeared to be more receptive to the explicitly strategy-oriented feedback from the more authoritative agent, i.e., the mentor, it may be especially fruitful to improve the mentor agent's feedback. We intend to continue analyzing the data from this and future studies in order to better understand how specific phrasing and different forms of metacognitive feedback affect student behavior. We have also been conducting studies to determine how to make the timing and content of strategy feedback more relevant to the student's current activities on the system.

In addition to analyzing and enhancing the agent feedback to promote metacognitive strategies and prepare students for future learning, we also plan to refine our HMM analysis technique. Enhanced HMM analysis could provide a better understanding of the different strategies employed by students when learning complex science topics and allow for more adaptive feedback suited to the current context of the students' activities. In particular, we intend to employ clustering of individual student HMMs to improve the accuracy of our HMM analysis and use sequence mining to pre-process the trace data in the HMM analysis to maintain more of the temporal information in the aggregated behaviors of HMM states.

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

References

- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education, 16*(2), 101–128.
- Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist, 40*(4), 199–209.
- Azevedo, R., & Witherspoon, A. M. (2009). Self-regulated use of hypermedia. In A. Graesser, J. Dunlosky, & D. Hacker (Eds.), *Handbook of metacognition in education*. Mahwah, NJ: Erlbaum.
- Azevedo, R., Witherspoon, A., Chauncey, A., Burkett, C., & Fike, A. (2009). MetaTutor: A metacognitive tool for enhancing self-regulated learning. *Annual Meeting of the American Association for Artificial Intelligence, Symposium on Metacognitive and Cognitive Educational Systems* (pp. 4–19).
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology-Enhanced Learning (RPTEL), 5*(2), 123–152.
- Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & Vanderbilt, T. T. A. G. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3), 363–392.
- Blair, K., Schwartz, D. L., Biswas, G., & Leelawong, K. (2007). Pedagogical agents for learning by teaching: teachable agents. *Educational Technology & Society: Special Issue on Pedagogical Agents, 47*(1).
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How people learn*. Washington, DC: National Academy Press.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61.
- Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393–451). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Burleson, W., Picard, R. W., Perlin, K., & Lippincott, J. (2004). A platform for affective agent research. *Workshop on Empathetic Agents, International Conference on Autonomous Agents and Multiagent Systems, New York, NY*.
- Chase, C. C., Chin, D. B., Opezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology, 18*(4), 334–352.
- D’Mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learners affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1), 45–80.
- D’Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems, 22*(4), 53–61.
- Forbus, K. D. (1984). Qualitative process theory. *Artificial intelligence, 24*(1–3), 85–168.
- Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning, 2*(2), 107–124.
- Hadwin, A. F., Winne, P. H., Stockley, D. B., Nesbit, J. C., & Woszczyna, C. (2001). Context moderates students’ self-reports about how they study. *Journal of Educational Psychology, 93*(3), 477–487.
- Jeong, H., & Biswas, G. (2008). Mining student behavior models in learning-by-teaching environments. *Proceedings of the First International Conference on Educational Data Mining, Montreal, Quebec, Canada* (pp. 127–136).
- Jonassen, D. H., & Ionas, I. G. (2008). Designing effective supports for causal reasoning. *Educational Technology Research and Development, 56*(3), 287–308.
- Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty’s Brain system. *International Journal of Artificial Intelligence in Education, 18*(3), 181–208.
- Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI ’97), Atlanta, GA* (pp. 359–366).
- Li, C., & Biswas, G. (2000). A Bayesian approach to temporal data clustering using hidden Markov models. *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 543–550).
- Li, C., & Biswas, G. (2002). A Bayesian approach for learning hidden Markov models from data. *Scientific Programming: Special Issue on Markov Chain and Hidden Markov Models, 10*(3), 201–219.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.
- Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students? Self-regulated engagements with computerized content. *Educational Psychology Review, 18*(3), 211–228.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*(1), 92–104.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice, 41*(4), 219–225.
- Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the

- process of conceptual change. *Review of Educational Research*, 63(2), 167–199.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and psychological measurement*, 53(3), 801–813.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.
- Roll, I., Alevin, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition: Applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, 2, 125–140.
- Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4), 534–574.
- 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.
- Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist*, 32(4), 195–208.
- Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (2007). Animations of thought: Interactivity in the teachable agent paradigm. In R. Lowe & W. Schnotz (Eds.), *Learning with animation: Research and implications for design* (pp. 114–140). Cambridge: Cambridge University Press.
- Schwartz, D. L., Chase, C., Chin, D. B., Oppezzo, M., Kwong, H., Okita, S. et al. (2009) Interactive metacognition: Monitoring and regulating a teachable agent. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 340–358). New York: Routledge Press.
- Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184.
- Tan, J., Biswas, G., & Schwartz, D. (2006). Feedback for metacognitive support in learning by teaching environments. *Proceedings of 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. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society, Nashville, TN* (pp. 695–700).
- Weinstein, C. E., Schulte, A. C., and Palmer, D. R. (1987). *The learning and study strategies inventory*. Clearwater, FL: H & H Publishing.
- Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). New York, NY: Taylor & Francis.
- Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, 2(2), 173–201.
- Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Erlbaum.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166–183.
- Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663–676.
- Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, 23(4), 614–628.