Wong, L.-H. et al. (Eds.) (2013). Proceedings of the 21st International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

How do students' learning behaviors evolve in Scaffolded Open-Ended Learning Environments?

Gautam BISWAS*, John S. KINNEBREW & Daniel LC MACK

Dept. Of EECS/ISIS, Vanderbilt University, Nashville, TN, USA
*gautam.biswas@vanderbilt.edu

Abstract: Metacognition and self-regulation are important components for developing effective learning in the classroom and beyond, but novice learners often lack these skills. Betty's Brain, an open-ended computer-based learning environment, helps students develop metacognitive strategies as they learn science topics. In order to better understand and improve the effect of adaptive scaffolding on students' cognitive and metacognitive skills, we investigate students' activities in Betty's Brain from a study comparing different forms of adaptive scaffolding. We measure students' cognitive and metacognitive processes from students' action sequences by (i) interpreting and characterizing behavior patterns using a cognitive/metacognitive model of the task, (ii) mapping students' frequently observed cognitive and metacognitive process patterns back into their overall activity sequences and measuring their effectiveness, and (iii) employing a binning method with clustering and visualization techniques to characterize the temporal evolution of these processes. Our experimental studies illustrate that the effectiveness and temporal changes in students' behaviors were generally consistent with the scaffolding provided, suggesting that these metacognitive strategies can be taught to middle school students in computer-based learning environments.

Keywords: pedagogical agents, teachable agent, metacognition, scaffolding, measuring metacognition

1. *Introduction*

Our research group has developed Betty's Brain, a computer-based open-ended learning environment (OELE; Land, 2000), where middle school students learn science topics by teaching a virtual agent named Betty using a visual causal map (Leelawong & Biswas, 2008). As she is taught, Betty can answer questions, explain her answers, and when requested by the students take quizzes developed by Mr. Davis, a Mentor agent. Betty's quiz performance helps students assess Betty's, and, therefore, their own knowledge. This motivates them to learn more, so they can help Betty improve her quiz score.

More recently, we have directed our attention to how students develop metacognitive strategies that include information seeking, solution construction, and solution assessment as they learn the science topics while they teach Betty. Our approach utilizes trace methodologies derived from students' actions and activity patterns in the environment to infer aspects of their metacognitive abilities (Aleven et al, 2006; Azevedo, et al., 2012; Hadwin et al, 2007). This is based on a *metacognition as events* hypothesis, which theorizes that the use of metacognitive strategies manifests as continually unfolding events that can be inferred from learners' behaviors.

In this paper, we extend our previous work on using sequence mining methods to discover students' frequently-used behavior patterns from their activity sequences as they work in the Betty's Brain system (Kinnebrew & Biswas, 2012). In particular, we extend our techniques for analyzing students' action sequences by (i) interpreting and characterizing behavior patterns using a cognitive/metacognitive model of the task, (ii) mapping students' frequently observed cognitive and metacognitive process patterns back into their overall activity sequences, (iii) using metrics to evaluate the effectiveness of these processes, and (iv) employing a binning method to characterize the temporal evolution of these processes based on their occurrence over the course of learning activities.

One of our primary goals in this paper is to study and analyze how students' learning behaviors evolve as they use the Betty's Brain system. We have developed a combination of sequence mining techniques (Kinnebrew, Loretz, & Biswas, 2013) and temporal interestingness measures (Kinnebrew, Mack, & Biswas, 2013) to capture and rank students' activity patterns. We employ a visual technique called heat maps (Wilkinson & Friendly, 2009) to study the temporal trends of the more interesting behavior patterns, and compare the trends across groups of students. The results in this paper and others (e.g., Biswas, Kinnebrew, & Segedy, 2013) represent a post hoc analysis of student behaviors, and it helps us evaluate the effectiveness of the feedback that Mr. Davis provides students to help them become more effective learners. Our longer term goal is to use the results of these analyses to track students' cognitive and metacognitive processes and measure the proficiency of their use as they work on their learning and problem-solving tasks. These results will help us develop better adaptive scaffolding mechanisms to support student learning.

2. Background: Metacognition

Metacognition is described as being made up of two constituent parts (Flavell et al, 1985; Veenman, 2012): (1) *Metacognitive knowledge*, which deals with awareness in the form of interplay between knowledge of one's abilities to perform tasks, the nature of the task, and the strategies one can employ to successfully perform the task; and (2) *Metacognitive control*, which includes activities related to goal selection, planning, monitoring, and evaluating one's cognitive processes in order to better regulate those processes in the future. Researchers have established strong links between learners' metacognitive abilities and their effectiveness in executing cognitive processes. For example, Winne (1996) characterizes cognition as dealing with knowledge of objects and operations on objects (the object level). Metacognition, on the other hand, corresponds to the meta-level that contains information about cognitive processes. Metacognitive monitoring brings the two levels together, as it describes the process of observing one's own execution of cognitive processes at the object level and exerting control over the object level using metacognitive knowledge and strategies.

An important implication of the interplay between cognition and metacognition relates to the dependence of metacognition on cognition (Land, 2000). In other words, metacognitive knowledge may not be sufficient for achieving success in learning and problem solving, especially for learners who lack the cognitive skills and background knowledge necessary for interpreting, understanding, and organizing critical aspects of the task at hand (Bransford, Brown, & Cocking, 2000). Learners may also lack knowledge of effective strategies (e.g., the ability to extract relevant information when reading a science text), and, therefore, resort to suboptimal strategies in performing their tasks (Azevedo, 2005; Kinnebrew & Biswas, 2012). Poor self-judgment abilities result in difficulties for monitoring and evaluating one's own effectiveness and progress, which can be a significant stumbling block in selecting and implementing relevant strategies in a timely manner. However, research studies have shown that with proper scaffolding, middle school students can improve their metacognitive awareness and develop effective metacognitive strategies (Kramarski & Mevarech, 2003). Recent results in Segedy, et al. (2013) also demonstrate that providing students' support in becoming more proficient in the cognitive processes when they need it, improves their overall effectiveness in the Betty's Brain system. Furthermore, we believe that developing the relevant cognitive skills will help students interpret and apply metacognitive strategies in a more effective manner.

3. Betty's Brain

Betty's Brain (Figure 1) is an open-ended learning environment (Land, 2000) that provides students with a learning context and a set of tools for pursuing authentic and complex learning tasks. Students explicitly teach Betty by constructing a causal map. For example, they may draw a causal link between *garbage and landfills* and *methane* to represent the relationship *garbage and landfills increase methane* (a greenhouse gas). Students can check what Betty knows by asking her questions, e.g., *if garbage and landfills decrease, what effect does it have on polar sea ice*? To answer questions, Betty uses qualitative reasoning that operates through chains of links from the source concept to the target concept (Leelawong & Biswas, 2008). The learner can further probe Betty's understanding by asking her to explain her answers. Betty illustrates her reasoning by explaining her thinking and highlighting concepts and links on the map as she mentions them. The goal for students using Betty's Brain is to teach Betty a causal map, whose correctness is determined in relation to a hidden, expert causal map.

Therefore, students' learning and teaching tasks are organized around three activities: (1) *reading* hypertext resources *to learn* the domain material, (2) *building and refining* a causal map, which

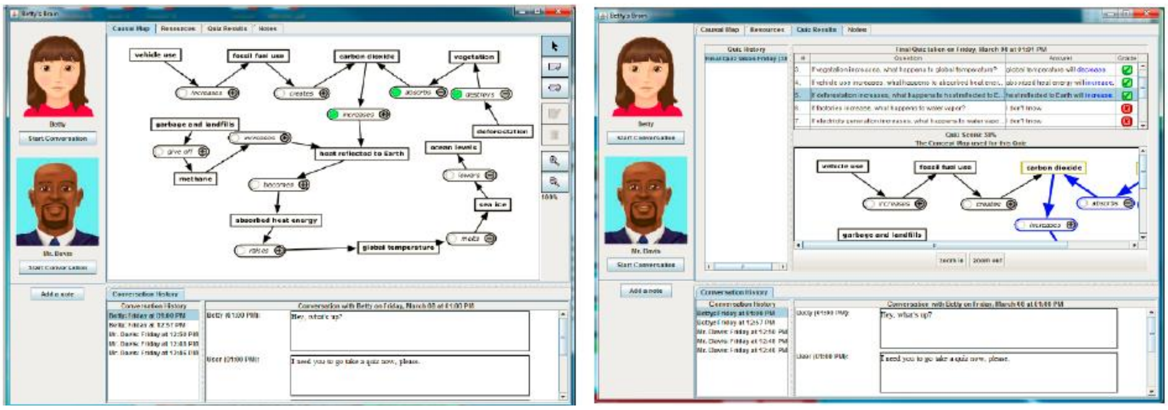

Figure 1: Betty’s Brain: (a) Map Editing Interface; (b) Quiz interface

captures the domain model, and (3) *asking* Betty *to take quizzes*. Students explicitly teach Betty by constructing a causal map. The map building and quiz interfaces for the system are shown in Figure 1.

Learners can assess their progress by having Betty take a quiz on one or all of the sub-topics that make up the causal map. Quiz questions are selected dynamically to reflect the current state of the student’s map; questions are chosen (in proportion to the completeness of the map) for which Betty will generate correct answers. The remaining questions produce incorrect or incomplete answers and they direct the student’s attention to erroneous and missing links, respectively. After Betty takes a quiz, her results, including the causal map she used to answer the questions appear on the screen as shown in Figure 1(b). The quiz questions, Betty’s answer, and the Mentor’s assigned grade, i.e., *correct*, *correct but incomplete*, or *incorrect* appear on the top of the window. Clicking on a question will highlight the causal links that Betty used to answer that question. To help students keep track of correct and incorrect links, the system allows students to annotate them with a green check-mark (correct), a red X (incorrect), or a gray question mark (not sure).

3.1 Cognitive/Metacognitive Process Model

We have developed a systematic approach to interpret students learning behaviors on the system. The model takes into account the close connection between the cognitive and metacognitive processes, and represents our expectations of the skills and strategies students need to develop to address the learning task effectively. Overall, this model, shown in Figure 2, includes three primary processes that students are expected to engage in while using Betty’s Brain: (1) *information seeking*, i.e., determining when and how to locate needed information in the resources, (2) *solution construction*, i.e., organizing one’s

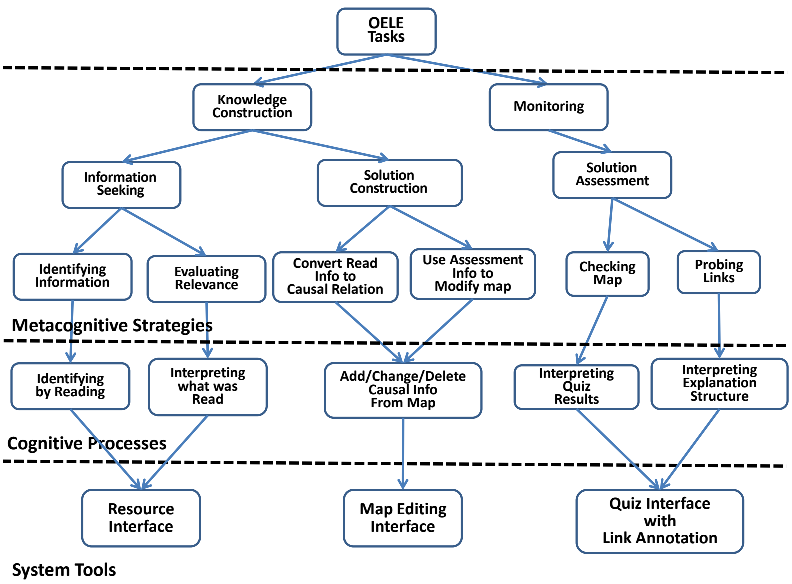

Figure 2: Cognitive/Metacognitive Task Model

developing understanding of the domain knowledge into structural components (e.g., causal links), and (3) *solution assessment*, i.e., assessing the correctness of one's causal model. (1) and (2) together represent strategies linked to *Knowledge Construction* (KC), whereas (3) pertains to strategies related to *Monitoring* (Mon). In executing metacognitive strategies, learners have to correctly execute related cognitive processes to be successful. Identifying information, for example, requires students to locate sentences that contain causal information as they read the resources and make sense of the content. Similarly, solution construction includes strategies for converting the acquired information into causal links and adding them to the appropriate place in the causal map.

Solution Assessment includes strategies for (1) checking the causal map, i.e., assessing the correctness of all or a part of the causal model, and (2) interpreting explanation structure, i.e., using the explanations to explicitly identify and annotate parts of the causal model as correct or incorrect. This makes it easier for the student to focus on parts of the map that need more work. Successful execution of monitoring metacognitive processes relies on students' abilities to execute cognitive processes for assessing the causal model (via questions, explanations, quizzes, and question evaluations) and recording progress (via note taking and annotating links with correctness information). The cognitive and metacognitive process model provides a framework for interpreting students learning activities and behaviors (activity sequences) on the system.

3.2 Measuring Cognition and Metacognition

We have developed sequence mining methods for analyzing students' learning activity sequences and assessing their learning processes as they work in Betty's Brain (Kinnebrew, Loretz, & Biswas, 2013, Segedy, Biswas, & Sulcer, 2013). In this paper, we extend this analysis and demonstrate the use of information gain measures and visualization methods to show how student behaviors evolve during the course of the intervention. In particular, we are interested in studying (1) whether students are able to interpret and apply KC and Mon strategies, and (2) whether this results in students' suboptimal behaviors gradually evolving into the use of more optimal strategies as the intervention progresses.

To assess the effectiveness of students' activities we calculate four measures: (1) map edit effectiveness, (2) map edit support, (3) monitoring effectiveness, and (4) monitoring support. Map edit effectiveness is calculated as the percentage of causal link additions, removals, and modifications that improve Betty's causal map. Map edit support is defined as the percentage of causal map edits that are supported by previous reading of pages in the resources that discuss the concepts connected by the manipulated causal link. Monitoring effectiveness is calculated as the percentage of question evaluations, quizzes, and explanations that generate specific correctness information about one or more causal links. For example, when Betty answers a quiz question correctly, all of the links she used to generate her answer are also correct. When students view the links Betty used to answer a correct question, they generate correctness information for each link (i.e., each link Betty used to answer the question is correct). Finally, monitoring support is defined as the percentage of causal link annotations that are supported by previous quiz questions and explanations. For support metrics, a further constraint is added: an action can only support another action if both actions occur within the same time window, and we calculated support for a ten minute time window.

The information for calculating the measures and deriving student behavior using sequence mining is extracted from students' activities on the system collected in log files. For example, if a student accesses a page in the resources, this is logged as a Read action that includes additional information, e.g., the page accessed. In this work, students' activity sequences are defined by six categories of actions: (1) Read, (2) Link Edit, (3) Query, (4) Quiz, (5) Explanation, and (6) Link Annotation. Actions are further distinguished by contextual details, such as the subtopic associated with the link, and the correctness of the link edited. Sequence mining techniques are applied to discover frequent behavior patterns for students in a given group are described in detail elsewhere, and not discussed in this paper (Biswas, Kinnebrew, & Segedy, 2013; Kinnebrew, Loretz, & Biswas, 2013; Kinnebrew & Biswas, 2012).

4. Method

We present post-hoc analyses from a 2012 classroom study with Betty's Brain in which 7th grade students in a middle Tennessee school learned about the greenhouse effect and climate change. The analysis that follows tests the effectiveness of two sets support modules designed to scaffold students' understanding of cognitive skills and metacognitive strategies important for success in Betty's Brain: (1) Knowledge Construction (KC) and Monitoring (Mon). The KC module provided support for information seeking and solution construction, and the Mon module helped students understand how to use Betty's quizzes to identify correct and incorrect links on the causal map. Participants were divided into three treatment groups. The KC group (KC-G) used a version of Betty's Brain that included the KC support module and a causal link tutorial that they could access at any time in the system. The tutorial allowed students to practice identifying causal relations in short text passages. The Mon group (Mon-G) used a version of Betty's Brain that included the Mon support module and a marking links correct tutorial that they could access at any time in the system. The tutorial presented practice problems in which students used the results of graded quiz questions and the related causal map to identify the links that could be marked as correct. Finally, the control group (Con-G) used a version of Betty's Brain that included neither the tutorials nor the support modules.

The KC module was activated when three out of a student's last five map edits were incorrect, at which point Mr. Davis would begin suggesting strategies for identifying causal links during reading. Should students continue to make incorrect map edits despite this feedback, the KC module activated a second tier of support: guided practice. During guided practice, students were moved to the causal link tutorial where they read short text passages and expressed the primary idea in the passage as a causal relation. When they worked on the tutorial, students were not permitted to access any other portion of the program. Students completed the tutorial session once they solved five problems correctly without making a mistake.

The Mon module was activated after the third time students did not use evidence from quizzes and explanations to annotate links on their map. At this time, Mr. Davis began suggesting strategies for using quizzes and explanations to identify and keep track of which links were correct. Additionally, Mr. Davis discouraged students from annotating links as being correct without using the suggested strategies. Should students continue to use quizzes and explanations without annotating links correctly, the Mon module provided students with guided practice. Like the KC tutorial, students had to complete five problems correctly on the first try to complete the tutorial session.

Seventy-three students, taught by the same teacher, participated in the study. Because use of Betty's Brain relies on students' ability to independently read and understand the resources, the system is not suited to students with limited English proficiency or cognitive-behavioral problems. Therefore, data from English as a Second Language (ESL) and special education students were not analyzed. Similarly, we excluded the data of students who missed more than two class periods of work on the system. Our experimental analysis used data collected from 52 students who participated in the study, with 15 students in the Con-G condition, 20 students in the KC-G condition, and 17 students in the Mon-G condition.

Learning was assessed using a pre-posttest design. Each written test consisted of five questions that asked students to consider a given scenario and explain its causal impact on climate change. Scoring was based on the causal relations that students used to explain their answers to the questions, which were then compared to the chain of causal relations used to derive the answer from the expert map. One point was awarded for each causal relationship in the student's answer that came from or was closely related to an expert causal link. The maximum combined score for the five questions was 16. Two coders independently scored a subset of the pre- and post-tests with at least 85% agreement, at which point the coders split the remaining tests and individually coded the answers and computed the scores.

Performance on the system was assessed by calculating a score for the causal map that students created while teaching Betty. This score was computed as the number of correct links (the links in the student's map that appeared in the expert map) minus the number of incorrect links in the student's final map. We also used the log data collected from the system to derive students' behavior patterns, interpret them using our cognitive/metacognitive model, and study the temporal evolution of the observed KC and Mon strategies over the period of the intervention.

5. *Results*

Repeated measures ANOVAs performed on the pre- to post-test gains revealed a significant effect of time on test scores ($F=28.66$, $p < 0.001$). Pairwise comparisons of the three groups revealed that the

Mon-G (learning gain mean (m) = 2.41, standard deviation (sd) = 1.92) had marginally better learning gains than KC-G (m = 1.28, sd = 2.33), which had better learning gains than the Con-G students (m = 1.03, sd = 1.99). The Mon-G learning gains were better than the Con-G gains ($p < .075$), indicating the two interventions may have resulted in better understanding of the science content. There was virtually no difference in the final map scores between the three groups. The small sample size and the large variations in performance within groups made it difficult to achieve statistical significance in these results. However, one positive aspect of this finding is that while students in the Mon-G and KC-G spent an average of 10% and 17% of their time in guided practice, respectively, they learned, on average, just as much, if not more, than the Con-G students.

We assessed students' overall behaviors using the effectiveness and support measures reported in section 3.2. The results in Table 1 show that the KC-G students had the highest scores on both map editing effectiveness and support, suggesting that the KC feedback did help students more effectively and systematically read and construct their causal maps (however, only the map edit support showed a statistically significant difference, KC-G > Con-G, $p = 0.02$, and the map edit effectiveness illustrated a trend, KC-G > Con-G, $p = 0.08$). However, the monitoring feedback did not help the Mon-G students do better than the other two groups for monitoring effectiveness or support. The Mon-G students did have the highest monitoring effectiveness, but the differences were not statistically significant. The Con-G students had the highest monitoring support average ($p < 0.10$, when comparing with other groups). It is not clear why the Mon or KC support and tutorials resulted in students performing less supported monitoring activities than the Con-G students.

Table 1: Effectiveness & Support Measures ((mean (std dev)) by Group

Measure	Con-G	KC-G	Mon-G
Map edit effectiveness	0.46 (0.13)	0.52 (0.07)	0.50 (0.12)
Map edit support	0.43 (0.25)	0.64 (0.19)	0.55 (0.23)
Monitoring effectiveness	0.30 (0.22)	0.32 (0.21)	0.40 (0.20)
Monitoring support	0.61 (0.30)	0.32 (0.40)	0.33 (0.32)

6. Temporal Evolution of Behaviors

We developed an approach for studying how students' learning behaviors and strategies evolve over time as the result of the scaffolds and feedback provided by the learning environment or the changing demands of the task over the course of learning. We have developed a methodology called the Temporal Interestingness of Patterns in Sequences (TIPS) along with a corresponding interestingness measure, for identifying and visualizing the most temporally-interesting patterns of student behavior (Kinnebrew, Mack, & Biswas, 2013).

To characterize the temporal evolution of pattern use, we sliced each student's sequence into a given number of bins and aggregated the pattern use frequencies for corresponding bins over a group of students using the following approach:

1. Map the patterns back to the individual student activity sequences with occurrence frequency over the length of the sequence. This is performed by slicing each sequence into n bins, such that each bin contains approximately ($100/n$) % of the student's actions in the sequence. Since corresponding bins for different students can be of different sizes (i.e., the number of actions in the bin) depending on the total number of actions the student performed, the frequency of a pattern in a bin is calculated as the number of occurrences divided by the size of the bin;
2. Compute the group frequency of the patterns for each bin as the average across all students in the group, and normalize the frequency counts as percentages to provide a standard basis for comparison across groups. The percentages allow direct comparison across groups in terms of the evolution of pattern use, even with different total frequencies for pattern use in the groups; and
3. Form a pattern vector for each group by considering each behavior pattern to be defined by a vector consisting of n feature values, where each feature value corresponds to the percentage frequency count for the corresponding bin.
4. Provide a ranking of the candidate patterns using Information gain (IG) as an interestingness measure applied to the temporal footprint of each pattern. IG is defined as the difference in

expected information entropy between one state and another state where some additional information is known (e.g., a set of data points considered as a homogeneous group versus one split into multiple groups based on the value of some other feature or attribute). In TIPS we apply IG to determine which patterns are the best descriptors of the data because knowledge of their occurrence provides the least amount of uncertainty about the temporal location of actions in the sequences. This IG measure for a pattern defines its temporal-interestingness in TIPS and is used to rank all candidate patterns in descending order, so the pattern that has the highest information gain will be ranked first.

5. For the highly-ranked patterns, visualize their temporal footprints using heat maps (Wilkinson & Friendly, 2009) to assess usage trends and spikes. Specifically, we employ a single-dimensional heat map where each temporal bin's value is its percentage of the total pattern occurrence. The heat map is generated by assigning a color to each bin, which is determined by where its value falls between the highest and lowest value in the heat map.

As an initial analysis of student learning behaviors, we ran the students' activity data through a sequence mining algorithm (Kinnebrew, Loretz, & Biswas, 2013). This algorithm identified 143 different action patterns that were observed in a majority of students. We ran our binning method on all 143 patterns to divide up the activity sequences into 5 bins, such that each bin contained about 20% of the students' actions in the learning environment. The TIPS algorithm (Kinnebrew, Mack, & Biswas, 2013) was then run on the 143 patterns, and the top four distinct behavior patterns ranked by IG are listed in Table 2. Patterns 1 and 2 can be classified more as KC strategies, and not surprisingly the KC group has the highest average frequency of use for these strategies. Similarly, patterns 3 and 4 involve interpretation and further probing into Quiz results, respectively, and, therefore, they may be classified as Mon strategies. Mon-G has a marginally greater frequency of use of these two strategies than the KC-G; however, the CON-G students show the highest average frequency of use for these strategies. When one looks at these results in conjunction with monitoring effectiveness, the MON-G students do have a higher value for that measure as compared to the other two groups. Surprisingly, however, the CON-G students had a much higher score than the Mon-G and KC-G students for the monitoring support measure. However, it is interesting to note that this did not translate to higher map scores and better pre-post test gains for this group.

Table 2: Comparison of Pattern Frequencies across Conditions for highest ranked TIPS patterns

Rank	Pattern	Avg. Frequency		
		CON	KC	MON
1	[Add incorrect link (AIL)] → [Remove incorrect link (RIL)] → [Read (R)]	1.4	5.0	0.9
2	[Read Multiple Pages(R-M)] → [Add incorrect link (AIL)]	1.7	5.0	1.9
3	[Add incorrect link (AIL)] → [Quiz (Q)] → [Remove incorrect link (RIL)]	5.3	4.2	4.3
4	[Quiz (Q)] → [Explanation (E)]	7.6	3.2	4.5

While these broad category differences in pattern use among students suggests that the expected effects on student learning behaviors for the two experimental conditions, the heat maps provide a visual comparison of the temporal evolution of individual activity patterns across conditions. Figure 3 illustrates the temporal evolution of the four behavior patterns listed in Table 2. In the heat maps, the lighter shades imply a lower frequency of occurrence for that pattern in a bin, and the darker shades imply higher frequencies of occurrence. The legend colored white denotes the lowest frequency of occurrence in that heat map expressed as a percentage of total occurrence for the corresponding group, whereas the legend shaded black denotes the highest percentage frequency of occurrence in the heat map.

Pattern 1: AIL $\rightarrow$ RIL $\rightarrow$ R, very likely represents a justifiably cautious strategy, where the student added a link then correctly surmised that the link was incorrect. It was not clear if this was a guess or the student used past information (e.g., a previous quiz result) to come to this conclusion. In any case, the student then decided to read the resources further, presumably to find the right link. The KC-G students used this pattern with much greater frequency that the Con-G and Mon-G students and the pattern use was uniformly distributed through the period of the intervention. For the Con-G students the use of this pattern increased monotonically through most of the intervention, but dramatically

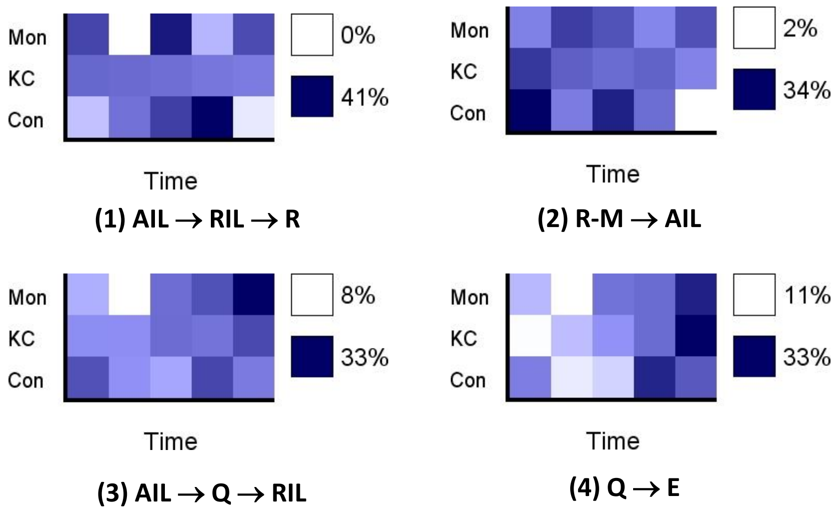

Figure 3: Temporal Evolution of Behavior Patterns

dropped off for the last bin. The Mon-G students use this pattern sporadically (their average frequency of use in the intervention was < 1), but the pattern's use peaked in the middle of the intervention.

Pattern 2: R–M → AIL, represents a suboptimal behavior, where students read multiple pages then added an incorrect link to the map. Though the KC-G students did this most often, Figure 3 shows that this behavior decreased in frequency for the KC-G students as the intervention progressed. The frequency of use of this pattern for Mon-G and Con-G groups was roughly the same. But the Con-G students' use of this pattern decreased with time, whereas for the Mon-G students the frequency of use was more uniform with a small increase at the end.

Pattern 3: AIL → Q → RIL represented a strategy where a student added an incorrect link and then took a quiz to see how the score would change. Depending on the outcome (in this case, the score likely decreased), the student determines that the link added was incorrect, and, therefore, deleted it. This generally represents a suboptimal trial-and-error strategy. Students in all three groups used this strategy, but the Mon-G and KC-G students used it less than the Con-G students. This may be attributable to the effectiveness of the Monitoring scaffolding. To study this pattern further we developed two pattern-specific measures: (1) a measure of cohesiveness of the pattern, i.e., in what percentage of the AIL → Q → RIL patterns was the delete action supported by the quiz result; and (2) a support measure, i.e., in what percentage of the AIL → Q → RIL patterns was the addition of the link supported by recent actions. The MON group had higher cohesiveness (41.9 to 38.0 and 37.3 for the CON and KC groups) and higher support (27.7 to 20.3 and 18.7 for the CON and KC groups) measures, implying that when they employed this pattern, they did so in a more systematic way than the other two groups.

Pattern 4: Q → E represents a good monitoring strategy, where the students use the explanation feature to identify the links Betty used in answering a question that appeared on the quiz. The Mon-G feedback provided by Mr. Davis emphasized the use of this approach to identify the correct versus incorrect links in their maps. The feedback had a positive effect. The Mon-G students used this feedback more frequently than the KC-G students, and the frequency of use increased as the intervention progressed for these groups. This can be attributed to the fact that the explanation feature becomes more useful as the size of the causal map grows. However, it was surprising that the Con-G students used the explanation feature with higher frequency than the other two groups, but their use of this behavior occurred more at the beginning and end of the intervention, with very little use in between. Perhaps, feedback from the Mentor would have led to greater use, and better map scores for the Con-G students.

7. Discussion and Conclusions

The results presented in Sections 5 and 6 provide evidence that a combination of theory-driven measures and data-driven mining techniques can be successfully employed to produce a more complete description of the type of metacognitive strategies students use and their effectiveness in their learning and problem-solving tasks. Furthermore, the temporal interestingness measures and the heat map visualization help us study how these behaviors evolve during the intervention. In this work on investigating cognitive and metacognitive processes in Betty's Brain, we had to carefully instrument the system to collect rich data on the students' activities and the context associated with those activities. Post hoc mining and analysis of this data revealed a number of interesting results. More generally, and perhaps most important, the results show (i) that it is possible to infer aspects of students' use of strategies through these data mining and analysis techniques combined with a cognitive/metacognitive model of the task, and (ii) that tracking student performance and related context information with respect to their activities allows us to better characterize how students' use of these strategies evolve, and how effective is the scaffolding provided by the system.

Our analyses in this study focused on students' information seeking, solution construction, and solution assessment strategies. Knowledge construction (KC) strategies include seeking out information, thinking deeply about the material to develop a sufficient understanding to apply it to model-building and problem-solving tasks. In particular, information structuring strategies in Betty's Brain help students with their map-building activities, which include understanding the structure of the causal model, the ability to construct it in parts, the ability to add links correctly to an existing structure, and also the ability to reason (e.g., answer questions, formulate hypotheses) with the evolving structure. The primary monitoring strategies relate to determining when and how to check the correctness of the current causal map, and then, in more detail, using the quiz (assessment) results to determine the correctness of individual links and identify parts of the map that are incomplete or still need work.

In summary, the analysis presented in this paper successfully employed our metacognition measurement framework to evaluate the effects of scaffolding support for metacognitive and cognitive processes important for success in Betty's Brain. Comparison of different versions of Betty's Brain, a version that provided very little scaffolding and no guided practice versus two experimental conditions: one that provided Knowledge Construction scaffolds and a second that provided Monitoring scaffolds, produced interesting results. Overall, the interventions resulted in changes in student behavior that were consistent with the provided scaffolding, implying that these metacognitive strategies can be taught and supported for middle school students in a computer-based learning environments. In more detail, the KC group tended to use KC strategies more often than the other two groups, but this was not true for students in the Mon group, who did not use monitoring strategies more often than the other two groups (see Table 2 and Figure 3). Furthermore, when we analyzed the temporal evolution of two monitoring-related behaviors, the positive result was that the Mon-G students applied the Q $\rightarrow$ AIL $\rightarrow$ Q pattern more effectively than a trial and error strategy as compared to the other two groups. Their use of this behavior pattern only increased toward the end, when the map building task became more difficult. In subsequent versions of the system, we may monitor the use of this behavior more closely, and provide students help with alternate strategies that are likely to be more effective in helping them complete their causal map. There were less distinct differences in the use of explanations after a quiz. All students used the explanation feature more often toward the end. To promote more uniform and effective use of explanations throughout the interventions, we will develop new feedback that better explain how to interpret quiz results and the role of explanations in probing the correctness of links used to answer questions (e.g., how to use the results from multiple quiz questions).

Further improvements in the scaffolding and feedback on solution construction and solution assessment strategies provided by the environment and creating repeated opportunities for students to practice them should lead to better science learning performance. An interesting implication of this work is that solution assessment strategies can be the key to better learning performance as well as prove to be the catalyst for developing more effective information seeking and solution construction strategies. The results presented in this paper are promising, but further analysis and more systematic experiments will have to be conducted to achieve conclusive results.

Acknowledgements

This work has been supported by NSF-IIS Award #0904387 and IES Award #R305A120186.

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., et al.. (2012) The Effectiveness of Pedagogical Agents' Prompting and Feedback in Facilitating Co-adapted Learning with MetaTutor. In *Intelligent Tutoring Systems* (pp. 212-221). Springer Berlin/Heidelberg.

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.

Biswas, G., Kinnebrew, J.S., & Segedy, J.R. (2013). Analyzing Students' Metacognitive Strategies in Open-Ended Learning Environments. In *Proceedings of the 35th annual meeting of the Cognitive Science Society*. Berlin, Germany, August.

Bransford J, Brown A, & Cocking R (eds) (2000) *How people learn*. National Academy Press Washington, DC, Washington.

Chi M, Glaser R, & Farr M (1988) *The nature of expertise*. Lawrence Erlbaum Associates, Inc.

Flavell J, Miller P, Miller S (1985) Cognitive de-velopment. Prentice-Hall Englewood Cliffs, NJ

Hadwin A, Nesbit J, Jamieson-Noel D, Code J, & Winne P (2007) Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2(2):107-124.

Kinnebrew J.S.& Biswas G. (2012) Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. In: *Proc. 5th International Conference on Educational Data Mining* (EDM 2012), Chania, Greece, 57-64.

Kinnebrew J.S., Loretz K.M., & Biswas G. (2013) A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, 5(1), 190-219..

Kinnebrew, J.S., Mack, D.L.C., & Biswas, G. (2013). Mining Temporally-Interesting Learning Behavior Patterns. In *Proceedings of the 6th International Conference on Educational Data Mining*. Memphis, TN, USA, 252-255.

Kramarski B. & Mevarech Z. (2003) Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal* 40(1):281-310.

Land S. (2000) Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development*, 48(3):61-78.

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.

Rowe, J., Shores, L., Mott, B., & Lester, J. (2011) Integrating learning, problem solving, and engagement in narrative-centered learning environments. *Intl Jour of Artificial Intelligence in Education*, 21(1-2):115–133.

Segedy, J.R., Biswas, G., & Sulcer, B. (2013). A Model-Based Behavior Analysis Approach for Open-Ended Environments. *Journal of Educational Technology & Society*. In Press.

Segedy, J.R., Biswas, G., Blackstock, E.F., & Jenkins, A. (2013). Guided Skill Practice as an Adaptive Scaffolding Strategy in Open-Ended Learning Environments. In Proceedings of *The 16th International Conference on Artificial Intelligence in Education*. Memphis, TN, USA, 532-541.

VanLehn K (1996) Cognitive skill acquisition. Annual review of psychology, 47(1):513-539.

Veenman M (2012) Metacognition in science education: Definitions, constituents, and their intricate relation with cognition. *Metacognition in Science Education*, pp. 21-36.

Wilkinson, L. & Friendly, M. (2009) The history of the cluster heat map. *The American Statistician*, 63(2):179-184.

Winne P (1996) A metacognitive view of individual differences in self-regulate learning. *Learning and individual differences*, 8(4):327-353.