Examining Self-Regulated Learning
in a Narrative-Centered Learning Environment: An
Inductive Approach to Modeling Metacognitive
Monitoring

Scott W. McQuiggan¹, Kristin L. Hoffman², John L. Nietfeld³,
Jennifer L. Robison¹, and James C. Lester¹

¹Department of Computer Science, North Carolina State University, Raleigh, NC
²Department of Curriculum and Instruction, North Carolina State University, Raleigh, NC
³Department of Instructional Technology, North Carolina State University, Raleigh, NC
{swmcquig, klhoffma, john_nietfeld, jlroboio, lester}@ncsu.edu

Abstract. Recent years have seen a growing interest in the development of self-regulated learning (SRL) models in computer-based environments and intelligent tutoring systems. In this paper, we explore SRL by investigating student monitoring, just one of many SRL components. We utilize an inductive approach to automatically construct models of student monitoring judgments from observations of student behavior in a narrative-centered learning environment, CRYSSTAL ISLAND. The resulting models can be used at runtime to accurately predict student monitoring and inform pedagogical decision-making. We believe that this approach may be applicable to modeling a variety of SRL constructs to create a suite of data-driven SRL models, which can inform pedagogical feedback, problem selection, levels of challenge, among many other scaffolding techniques.

1 Introduction

Narrative is the subject of increasing interest within the intelligent tutoring systems community. Recently, researchers have begun to develop narrative-centered learning environments (NLEs) that combine story contexts and pedagogical support strategies to deliver effective, engaging educational experiences. Contextualizing learning within narrative affords the use of artificial intelligence techniques that tailor narrative and educational content to students’ actions, affective states, and abilities. Drawing on an interdisciplinary body of work, including intelligent tutoring systems, embodied conversational agents, and serious games, these environments offer the promise of adaptive, motivating learning experiences to students. NLEs are currently under investigation in a range of domains, including military soft-skills training [11, 23], anti-bullying education [1], health intervention education [15], and science learning in microbiology and genetics [18].

By incorporating learning into narrative-based, virtual environments, investigators hope to tap into students’ innate facilities for crafting and understanding stories. Contextualized learning experiences are known to encourage regulated learning
behavior [21]. Narrative is also well suited to alternative learning paradigms such as guided discovery and inquiry-based learning. Leveraging stories’ ability to draw audiences into plots and settings, NLEs can introduce novel perceptual, emotional, and motivational experiences, as well as establish connections between narrative content and pedagogical subject matter in young learners [27]. Further, NLEs can effectively support the factors shown to contribute to student levels of motivation [14]. Such contextual experiences influence student learning and motivation [13] by providing an environment whereby students are self-directed in their goal-based behaviors and yet orientation toward the goals themselves can be guided by the organization and structure of the learning environment [28]. NLEs allow students a degree of autonomy and control within a structured context that is particularly fertile for the development of self-regulation.

Understanding how learners develop and improve skills of self-regulation in computer-based environments is becoming increasingly important. Self-regulated learning (SRL) refers to learning that results from students’ self-generated thoughts and behaviors that are systematically oriented toward the attainment of their learning goals [30]. Research in SRL has made significant advancements as models have been developed and refined [3, 9, 22, 33, 35]. However, there is now a shift in efforts to understand self-regulation not only in traditional learning environments but also in computer-based environments capable of providing intelligent tutoring [2, 8, 25]. This poses significant challenges for research designs because such environments are very complex and may require additional processing demands from the user [12, 26]. In order to provide sophisticated models of the development of SRL in computer-based environments (i.e., NLEs) and subsequently build intelligent tutoring systems that support the development of SRL we must delve into the models and start to examine the relationships between the key variables within the models. Models of SRL, at the broadest levels, are composed of strategic, metacognitive, and motivational components [35]. In this work we begin to examine the function of variables within these components by students working within a rich NLE.

Self-regulated learners continuously monitor cognition, motivation, and behavior guided by goals and the constraints of the environment [22]. Metacognitive monitoring skills and the regulation of strategies and tactics are core components within information processing models of self-regulation [3] and the development of human expertise in general [7]. More accurate monitoring has been shown to lead to improved self-regulation that, in turn, translates into improved performance [31]. Thus, intelligent tutoring systems would be able to further tailor instruction if they were able to effectively and accurately diagnosis student monitoring. Traditionally this is accomplished through student reports of monitoring judgments. However, this paper reports on the results of a study that investigates an inductive approach to modeling metacognitive monitoring in a narrative-centered learning environment, CRYSTAL ISLAND. The results of this experiment indicate that we can build accurate computational models of student monitoring that can operate efficiently at runtime to diagnose student monitoring without the use of interruptive self-reporting.
2 Crystal Island

CRYSTAL ISLAND is a narrative-centered learning environment built on Valve Software’s Source™ engine, the 3D game platform for Half-Life 2. CRYSTAL ISLAND features a science mystery set on a recently discovered volcanic island. The curriculum underlying CRYSTAL ISLAND’s science mystery is derived from the North Carolina state standard course of study for eighth-grade microbiology. Students play the role of the protagonist, Alyx, who is attempting to discover the identity and source of an infectious disease plaguing a newly established research station. The story opens by introducing the student to the island and members of the research team for which the protagonist’s father serves as the lead scientist. Several of the team’s members have fallen gravely ill, including Alyx’s father. Tensions have run high on the island, and one of the team members suddenly accuses another of having poisoned the other researchers. It is the student’s task to discover the outbreak’s cause and source and either acquit or incriminate the accused team member.

CRYSTAL ISLAND’s setting includes a beach area with docks, a large outdoor field laboratory, underground caves, and a research camp. Throughout the mystery, the student is free to explore the world and interact with other characters while forming questions, generating hypotheses, collecting data, and testing hypotheses. The student can pick up and manipulate objects, take notes, view posters, operate lab equipment, and talk with non-player characters to gather clues about the source of the disease. During the course of solving the mystery, the student is minimally guided through five problems corresponding to the North Carolina standard course of study for the eighth-grade microbiology curriculum. The first two problems focus on pathogens, including viruses, bacteria, fungi, and parasites. The student gathers information by interacting with in-game pathogen “experts” and viewing books and posters in the environment. In the third problem, the student is asked to compare and contrast her knowledge of four types of pathogens. In the fourth problem, the student is guided through an inquiry-based hypothesis-test-and-retest problem. In this problem she must complete a “fact sheet” with information pertaining to the disease afflicting members of the CRYSTAL ISLAND research team. Once the “fact sheet” is completed
and verified by the camp nurse, the student completes the final problem concerning the treatment of viruses, bacteria, fungi, and parasites, and selects the appropriate treatment plan for sickened CRYSTAL ISLAND researchers. The story and curriculum are interwoven throughout the interactive experience.

3 Method

3.1 Participants

The participants included 59 eighth grade students from a highly diverse (e.g., 46% minority; 32% free/reduced lunch) magnet school in Raleigh, North Carolina. Thirty-two were boys and 27 were girls. The students ranged in age from 13 to 15 ($M = 13.73$, $SD = 0.59$). Approximately 56% of the student participants were Caucasian ($n = 33$), 39% were African-American ($n = 23$), and 5% were Hispanic or Latino ($n = 3$). The students participated as part of their science class.

3.2 Materials and Apparatus

Materials consisted of the following computer-based instruments:

- **AGQ (Achievement Goals Questionnaire):** This is a 12-item scale that measures achievement goal orientation in the form of four factors: mastery approach, mastery avoidance, performance approach, and performance avoidance [5]. The four factors have been shown to have strong reliability and validity through a validation study [6]. The AGQ was given both as a pretest and posttest.

- **Goals Inventory** [24]. This 17-item inventory measures one's tendency to adopt mastery and performance goals and is based upon Dweck and Leggett's [4] seminal paper on this subject. The items were answered on a five-point Likert scale. Scores for the mastery goals variable could range from 12 to 60 and scores for the performance goals variable could range from five to 25. The Goals Inventory was given both as a pretest and posttest.

- **Gaming Survey.** The Gaming Survey was created for this study and asked questions about experience with video games and computer use. For this study we were particularly interested in two 5-point Likert scale items that included, “Do you play videogames,” that measured the frequency of game play and, “How skilled are you when playing video games,” that measured perceived skill in video game play. The Gaming Survey was given only as a pretest.

- **Situational Interest rating.** At the end of CRYSTAL ISLAND interaction participants were asked to rate their interest in the game (“Did you enjoy this game”) on a 10 point Likert scale with one being “Not at all true of me” and 10 being “Very true of me.” Participants rated their interest a second time following feedback from the Scoreboard.
4 Procedure

Before playing, all students were given background information on CRYSTAL ISLAND in addition to a sheet listing the characters of CRYSTAL ISLAND and a map of the island. The character handout includes images of the characters, their names, and their narrative roles (i.e., virus expert, or camp nurse). The map identifies the layout of the virtual environment and the spatial relationship between areas of interest, such as the dining hall, the research lab, and the infirmary. Students completed pretests and then played CRYSTAL ISLAND for 35 minutes. At an interval of 90 seconds during interaction with CRYSTAL ISLAND, participants were asked, “How well are you accomplishing your overall goal for the game?” They responded by entering a number from zero to 100 where zero corresponded to “not confident at all” and 100 corresponded to “very confident.” These monitoring judgments were collected through an in-game dialog box and recorded in the same log tracking student behavior in the CRYSTAL ISLAND environment. Following game play the participants completed the AGQ and Goals Inventory.

5 Results

5.1 Summary of Previous Results

Nietfeld et al. [20] examined what variables impacted performance in CRYSTAL ISLAND. Specifically, they investigated the relationship between several SRL variables of interest (goal orientation, monitoring, and situational interest) and CRYSTAL ISLAND outcome measures (number of actions completed, number of goals completed, number of mystery solution guesses, and score). Monitoring was clearly the most prominent SRL variable, it showed significant correlations with actions completed \( (r = .33) \), goals completed \( (r = .59) \), and score \( (r = .74) \). Monitoring also showed a significant negative correlation \( (r = -.45) \) with number of guesses indicating that students who were confident that they were attaining their goals tended to see less of a need to risk making a guess without having all of the necessary information. This decision appears to be a metacognitively accurate one in that the payoff was the tendency for higher game scores in the end. Neither the mastery or performance facets from the Goals Inventory nor the mastery approach or performance approach variables from the AGC showed significant relationships with the CRYSTAL ISLAND outcome variables. A similar lack of relationships held for the situational interest variable. The mastery avoid facet of the AGC did show negative relationships with goals completed \( (r = -.26) \) and score \( (r = -.39) \). In sum, these findings reveal the importance and centrality of monitoring with performance.
5.2 Modeling Metacognitive Monitoring

To accurately model metacognitive monitoring judgments we utilize procedures which have been used to model self-efficacy [17] and affect [16] using the WEKA machine learning toolkit [34]. Both a naïve Bayes classifier and decision tree model were learned. Naïve Bayes and decision tree classifiers are effective machine learning techniques for generating preliminary predictive models. Naïve Bayes classification approaches produce probability tables that can be incorporated into runtime systems and used to continually update probabilities for assessing student self-efficacy levels. Decision trees provide interpretable rules that support runtime decision making. Runtime systems with decision trees monitor the condition of the attributes in the rules to determine when conditions are met for assigning particular values of student self-efficacy. Both the naïve Bayes and decision tree machine learning classification techniques are useful for preliminary predictive model induction for large multidimensional data. Because it is unclear precisely which runtime variables are likely to be the most predictive, naïve Bayes and decision tree modeling provide useful analyses that can inform more expressive machine learning techniques (e.g., Bayesian networks) that also leverage domain experts’ knowledge.

A tenfold cross-validation analysis was utilized to obtain an estimate of model error. In this method the data is broken into ten equal partitions. In each of the ten iterations (folds), nine partitions are used to construct the model and one partition is held out for testing. Each fold uses a unique partition for testing. Tenfold cross-validation is widely used for obtaining a sufficient estimate of error [34].

Data consisted of traces of student behavior in the CRYSTAL ISLAND learning environment, including all actions, visited locations, goals accomplished, and character interactions. Values (0-100) of student monitoring judgments served as class labels.

The results (Table 1) indicate that the decision tree model is able to accurately predict student metacognitive monitoring judgment values with 93% accuracy. The decision tree model result is statistically significant from the performance of the naïve Bayes model ($\chi^2(1, N = 2752) = 1486.2, p < 0.0001$) and a baseline model ($\chi^2(1, N = 2752) = 1740.2, p < 0.0001$). Here, we define a baseline model which constantly predicts the most frequently occurring monitoring judgment value (e.g., 100). Surprisingly, the baseline model significantly outperforms the naïve Bayes model ($\chi^2(1, N = 2752) = 20.4, p < 0.01$).

Table 1. Modeling metacognitive monitoring results.

<table>
<thead>
<tr>
<th>Model</th>
<th>Correctly Classified Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>20.76%</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>14.16%</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>93.36%</td>
</tr>
</tbody>
</table>
6 Discussion

This study represents a first step in examining SRL variables in a rich and interactive NLE. A major implication of this study is the centrality that monitoring judgments have within SRL [3]. Future work should follow up not only on the monitoring judgments themselves in NLEs but also the level of calibration accuracy of such judgments, which have been shown in other contexts to be significant predictors of performance [19].

Further, we find that through observation of student behavior in the CRYSTAL ISLAND learning environment that we can sufficiently predict student monitoring values, utilizing a constructed decision tree model, at runtime. Future work should also examine whether similar approaches can be used to model levels of calibration accuracy in addition to other SRL variables. By combining models of student metacognitive monitoring with other models, such as self-efficacy [17], we may come to better understand student self-regulation in real-time affording adaptive pedagogical strategies tailored to individual students. Knowledge of how, when, and where students self-regulate via metacognitive and motivational monitoring judgments within a NLE may in turn aid teachers in developing instructional approaches that support self-regulated learning in both traditional and technology-oriented classrooms. If we are able to extend models of student metacognitive monitoring to understand student levels of calibration, we believe we would be able to further tailor pedagogical interventions for students who are over-confident and increase efficacy of under-confident students. In fact, pedagogical feedback on problem-solving performance has the greatest impact when students are confident but the solution is wrong [10]. On the other hand, similar feedback would be less useful for students with low confidence. Instead, when students are in low-confident situations, pedagogical strategies should include instruction and information [10].

Thus, the utility in modeling self-regulatory components, such as metacognitive monitoring, is the ability to better construct the content of pedagogical feedback.

With regard to limitations, the experiment was designed to control for time on task, allowing 35 minutes for the intervention. As a result of this constraint and the amount of content in CRYSTAL ISLAND, not all participants had finished solving the CRYSTAL ISLAND mystery at the end of the 35 minute session. An alternative design, which will be adopted in future work, will control for task completion rather than time on task.

7 Conclusion

Narrative is receiving increasing attention in the ITS community as a medium for contextualizing learning in meaningful ways while creating rich, engaging experiences for learners. NLEs provide a framework for learning engagement and self-regulation by providing a supportive context for student control and autonomy, necessary for the growth of self-regulated learning. Thus, it will be required by future intelligent NLEs to diagnose, understand, and respond to student self-regulation. One approach to beginning the exploration of developing intelligent SRL detection models...
is the use data-driven methodologies that can enrich runtime SRL models. The results of this study indicate that an inductive approach leads to models of metacognitive monitoring that are capable of accurate runtime diagnosis.

The results highlight several important directions for future work. First, analysis of student calibration in metacognitive monitoring accuracy should be investigated. This is a necessary extension to further customize pedagogical feedback on student metacognitive monitoring. Second, it is important to understand the misclassifications of induced models and how the misclassifications may inadvertently affect pedagogical strategies. Third, it is necessary to begin to develop and systematically evaluate how to scaffold student learning experiences in light of information obtained through induced SRL models, such as the models of metacognitive monitoring described in this work. Lastly, the inductive approach holds much promise for modeling a variety of SRL components (e.g., efficacy, affect, interest). Future work should investigate the merit of combining data-driven SRL models for scaffolding student learning experiences in narrative-centered learning environments and identifying key features of student behavior useful for diagnosing self-regulated learning.

Acknowledgments

The authors wish to thank Hiller Spires, Kim Turner, Betty Walsh, Ada Lopez, and members of the IntelliMedia lab for their assistance, Omer Sturlovich and Pavel Turzo for use of their 3D model libraries, and Valve Software for access to the Source™ engine and SDK. This research was supported by the National Science Foundation under Grant REC-0632450 and CNS-0540523. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References