Discovering Behavior Patterns of Self-Regulated Learners in an Inquiry-Based Learning Environment

Jennifer Sabourin, Bradford Mott, James Lester

North Carolina State University, Raleigh, North Carolina
{jlrobiso, bwmott, lester}@ncsu.edu

Abstract. Inquiry-based learning has been proposed as a natural and authentic way for students to engage with science. Inquiry-based learning environments typically require students to guide their own learning and inquiry processes as they gather data, make and test hypotheses and draw conclusions. Some students are highly self-regulated learners and are able to guide and monitor their own learning activities effectively. Unfortunately, many students lack these skills and are consequently less successful in open-ended, inquiry-based environments. This work examines differences in inquiry behavior patterns in an open-ended, game-based learning environment, CRYSTAL ISLAND. Differential sequence mining is used to identify meaningful behavior patterns utilized by Low, Medium, and High self-regulated learners. Results indicate that self-regulated learners engage in more effective problem solving behaviors and demonstrate different patterns of use of the provided cognitive tools. The identified patterns help provide further insight into the role of SRL in inquiry-based learning and inform future approaches for scaffolding.

Keywords: Self-regulation, inquiry-based learning, game-based learning.

1 Introduction

Inquiry-based learning has been the focus of recent attention in both traditional classrooms [1, 2] and intelligent tutoring systems [3–5]. Inquiry-based learning has achieved this popularity primarily due to its use of authentic problem-solving scenarios and because the student is put in control of her own learning. During this process, the student is expected to play an active part in "making observations, formulating hypotheses, gathering and analyzing data, and forming conclusions from that data" [5]. However, inquiry-based learning environments are naturally very open-ended and may provide little guidance to students on when and how to engage in these behaviors. Without sufficient guidance, students are less likely to learn effectively [1, 2].

To be successful in open-ended, inquiry-based environments students must be capable of setting meaningful learning objectives [6]. They must then identify activities, behaviors, and strategies that may achieve these learning goals, monitor and evaluate their progress and alter their behavior and strategies accordingly. Together these skills form the foundation of self-regulated learning. Self-regulated learning (SRL) can be described as "the process by which students activate and sustain cognitions, behav-

iors, and affects that are systematically directed toward the attainment of goals" [7]. Unfortunately, students can demonstrate a wide range of fluency in their SRL behaviors [8], with some students lagging behind their peers in their ability to appropriately set and monitor learning goals.

This work seeks to identify the patterns of inquiry behaviors characteristic of self-regulated learners during game-based learning. It investigates these behaviors in the context of the CRYSTAL ISLAND game-based learning environment. CRYSTAL ISLAND is an open-ended game for middle school science in which students engage in inquiry behaviors of gathering evidence, forming and testing hypothesis, and reporting conclusions. Students are classified as Low, Medium, or High self-regulated learners based on evidence of goal setting and monitoring behaviors. Differential sequence mining [9] techniques are used to identify patterns of behavior that occur at statistically different frequencies between the classes of self-regulated learners. Results suggest differences in how students use tools, monitor their progress, and draw conclusions based on relevant information. These findings suggest that self-regulated learners engage in fundamentally different types of inquiry behaviors and point to methods for supporting the inquiry of students who do not have strong SRL skills.

2 Background

The ability to set learning goals, identify successful strategies, and evaluate personal success is the hallmark of self-regulated learning. Students who exhibit self-regulated learning (SRL) skills are able to drive their own learning and are often more successful in learning tasks and academic settings [10]. While SRL skills can be taught and often improve with practice [11], students who have not yet developed appropriate SRL strategies are more likely to flounder in self-guided, inquiry-based learning environments [6]. However, there is evidence that with appropriate scaffolding, these environments can improve learning as well as aid in development of SRL and inquiry skills [5, 12, 13].

Consequently, identifying and scaffolding metacognitive behaviors such as self-regulated learning (SRL) in open-ended environments has been a focus of much work in the intelligent tutoring systems community. For example, in MetaTutor, a hypermedia environment for learning biology, think-aloud protocols have been used to examine which regulatory strategies students use, while analysis of students' navigation through the hypermedia environment helps to identify profiles of self-regulated learners [13]. Similarly, researchers have identified patterns of behavior in the Betty's Brain system that are indicative of self-regulation [14] and utilized sequence mining techniques to further explore successful learning patterns [9].

Prior work exploring self-regulated learning in CRYSTAL ISLAND has utilized evidence of goal setting and monitoring to distinguish Low, Medium, and High classifications of SRL tendencies [15]. Further analyses demonstrated that Medium and High SRL students have both higher prior knowledge and higher learning gains than Low SRL students. This suggests that Low SRL students start with some disadvantage and that the overall gap in knowledge is increased after interactions with CRYSTAL

ISLAND. Though all groups have significant learning gains, Low SRL students are not experiencing the same benefits of interaction with CRYSTAL ISLAND. Further analyses suggest that High SRL students may be making better use of the curricular resources in CRYSTAL ISLAND than Medium or Low SRL students. These findings have highlighted the need to better understand the inquiry behaviors of High self-regulated learners and how these patterns can be used to inform scaffolding of the Low SRL students.

3 Method

The investigation of SRL behaviors was conducted with students from two North Carolina middle schools interacting with CRYSTAL ISLAND, an open-ended game-based learning environment being developed for the domain of microbiology that is aligned with the North Carolina Standard Course of Study for eighth grade science [16].

3.1 CRYSTAL ISLAND

CRYSTAL ISLAND features a science mystery set on a recently discovered volcanic island. The student plays the role of a visitor who recently arrived on the island in order to see her sick father. However, the student gets drawn into a mission to save the entire research team from a spreading outbreak. The student explores the research camp from a first-person viewpoint and manipulates virtual objects, converses with characters, and uses lab equipment and other resources to solve the mystery. As the student investigates the mystery, she completes an in-game diagnosis worksheet in order to record findings, hypotheses, and a final diagnosis. This worksheet is designed

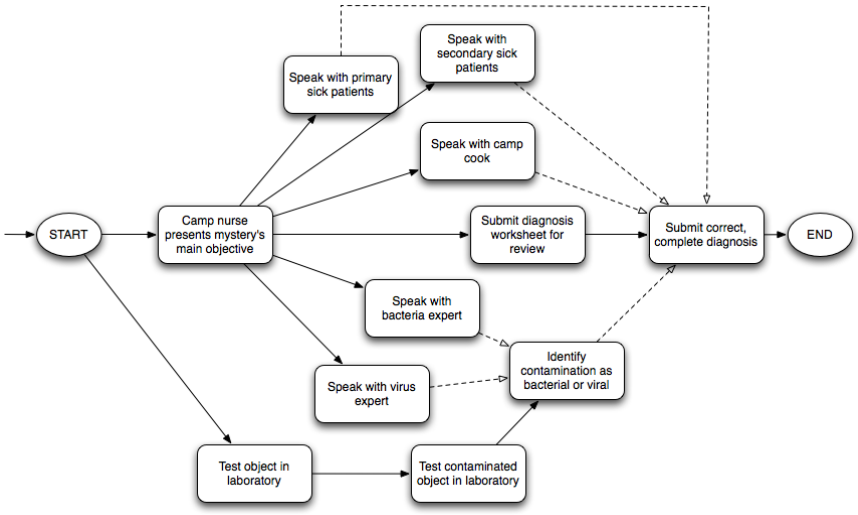

Fig. 1. Goal ordering in CRYSTAL ISLAND

to scaffold the student's problem-solving process and provide a space for the student to offload any findings gathered about the illness. The mystery is solved when the student submits a complete, correct diagnosis and treatment plan to the camp nurse.

To successfully complete the mystery, students must achieve several partially ordered goals (Figure 1). The goal topology indicates that many data-collection tasks are encouraged for students. They should converse with subject matter experts to learn about the underlying science content. They should discuss symptoms and possible sources of the outbreak with sick characters. They should read posters and books about different illnesses to help narrow down which diseases match the patients' symptoms. As students work towards solving the problem, they have two primary means to test their hypotheses. The first is through equipment in the camp's laboratory where students run tests on food objects to see if they are contaminated with pathogens, mutagens, or carcinogens. The second is through the diagnosis worksheet where they keep track of their hypothesized source and type of illness. This worksheet can be checked by the camp nurse for correctness.

While there is a subset of tasks that are strictly necessary to solve the mystery, there are a variety of tasks that are optional, but beneficial, for learning and problem-solving activities. For example, the diagnosis worksheet contains many fields to help students keep track of their hypotheses and thoughts, though only one small portion is required for reporting their final conclusions. Additionally, reading posters and books and talking with subject matter experts are helpful but not required to solve the mystery. Understanding how students choose to use these features of the learning environment is important for understanding effective inquiry strategies and how these strategies relate to self-regulated learning.

3.2 Study Procedure

A study with 450 eighth grade students interacting with the CRYSTAL ISLAND environment was conducted. After removing subjects with incomplete data or who experienced logging errors, there were 400 students remaining. Among the remaining students, there were 193 male and 207 female participants varying in age and ethnicity. Participants interacted with CRYSTAL ISLAND in their school classroom, although the study was not directly integrated into their regular classroom activities. Pre-study materials were completed during the week prior to interacting with CRYSTAL ISLAND. The pre-study materials included a demographic survey, researcher-generated CRYSTAL ISLAND curriculum test, and several personality questionnaires.

Immediately after solving the mystery, or after 55 minutes of interaction, students moved to a different room in order to complete several post-study questionnaires including the curriculum post-test. Students also completed two questionnaires aimed to measure students' interest and involvement with CRYSTAL ISLAND.

During the interaction students were prompted every seven minutes to self-report their current mood and status through an in-game smartphone device. Students selected one emotion from a set of seven options, which included the following: anxious, bored, confused, curious, excited, focused, and frustrated. After selecting an emotion, students were instructed to type a few words about their current status in the game,

similarly to how they might update their status in an online social network. These typed statements were tagged for evidence of goal setting and monitoring and used to classify students as High (*n=131*), Medium *(n=120*), or Low (*n=149*) SRL. (See [15] for more details.)

4 Identifying Behavior Patterns

Prior findings [15] on the differences in learning between Low, Medium, and High SRL students in CRYSTAL ISLAND prompted the current work to investigate differences in behavior patterns and inquiry strategies. Specifically, we sought to determine whether students interacted with CRYSTAL ISLAND in measurably different ways given their level of SRL skills. We also hoped to discover effective patterns utilized by High self-regulated learners that could be used to inform scaffolding for less skilled students. The exploratory nature of these questions and the desire to compare patterns across groups motivated the use of the differential sequence mining approach described by Kinnebrew et al. [9].

4.1 Action Abstraction

The first step to identify meaningful behavior patterns was to transform the highly detailed trace logs from interactions with CRYSTAL ISLAND into a more abstract representation of the overall behaviors being performed. This involved removing irrelevant or uninteresting actions (e.g., entering buildings, or manipulating individual objects) and grouping together instance of similar behaviors (e.g., reading a book on influenza and then a book on ebola).

In total, four general action types were identified as important distinguishing behaviors: TALK, READ, TEST, and WORKSHEET (Figure 2). The first two actions represent the primary source of gathering data in the environment, while the second two represent the primary problem-solving tasks and hypothesis testing tasks. These behaviors are central to the inquiry-based problem-solving in CRYSTAL ISLAND. Additional details were also considered for each action and are described below:

- **TALK:** One of the primary ways students gather information is through talking with in-game characters. Students may talk with patients to learn about the symptoms of their illness (TALK$_{SYM}$). There are also experts on pathogens, bacteria, and viruses that students may talk to (TALK$_{PATH}$, TALK$_{BAC}$, TALK$_{VIR}$). Finally, some of the characters also describe the nature of the illness and how it spread to students and provide details about the specific problem solving task (TALK$_{PROB}$).
- **READ:** There are several books and posters scattered around the environment that students may use for additional information. Many of these resources cover the same topics as conversations with experts on the island (READ$_{PATH}$, READ$_{BAC}$, READ$_{VIR}$). There is also a variety of books and posters that describe specific diseases (READ$_{DIS}$).

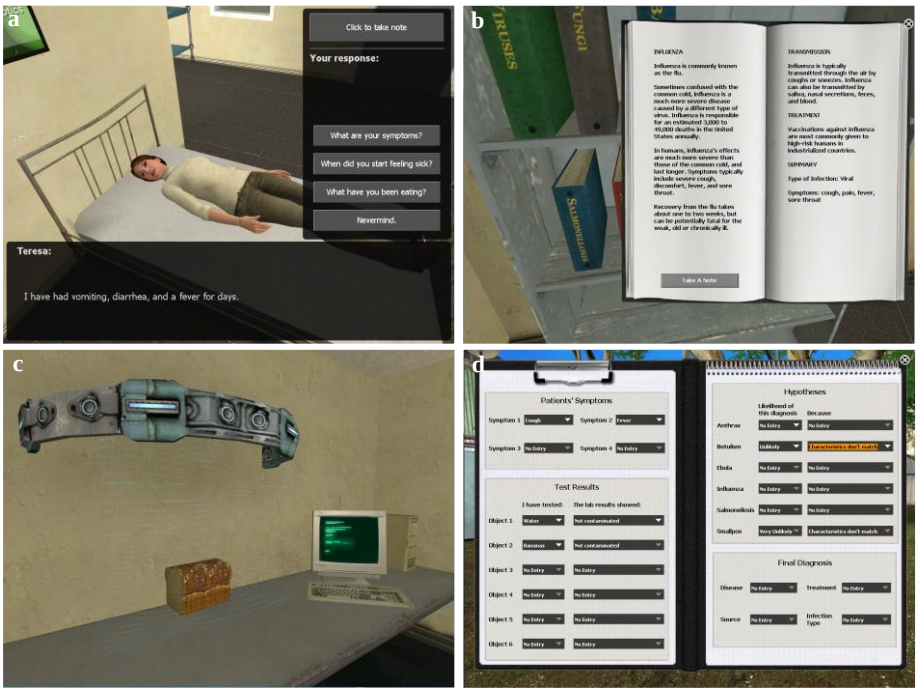

Fig. 2. Targeted behaviors (a) TALK, (b) READ, (c) TEST, (d) WORKSHEET

- **TEST:** To identify contaminated items students must run tests on individual food items. They must also specify whether they are testing the item for a pathogen, mutagen or carcinogen. Based on the nature of the illness, students should rule out mutagen or carcinogen as possible sources and testing for this is considered irrelevant (TEST$_{IRR}$). Tests for pathogens are identified as correct (TEST$_{CORR}$) if the proper food item was selected and incorrect (TEST$_{INC}$) otherwise.
- **WORKSHEET (WS):** The diagnosis worksheet is where students keep notes about their findings and hypotheses. There are several sections of information that can be filled out. They can record symptoms of patients (WS$_{SYM}$) and the results of their tests (WS$_{TEST}$). They can also keep track of hypotheses (WS$_{HYP}$) about individual diseases and their reasoning. The final section of the worksheet (WS$_{REP}$) is used to report their final conclusions to the nurse in order to complete the mystery.

4.2 Differential Sequence Mining

To identify patterns of behavior which were statistically different between Low, Medium, and High SRL students we utilized a differential sequence mining algorithm adapted from Kinnebrew et al. [9]. This approach identifies two metrics for representing the frequency of a pattern in different groups. The sequence support (*s-support*) metric refers to the percentage of sequences the pattern occurs in, regardless of frequency. Alternatively, the instance support *(i-support*) metric represents the average

number of times the pattern occurs per sequence. The primary adaption was to allow for comparison across the three groups where the original algorithm only compares between two populations. The adapted algorithm can be summarized in the following steps:

- *Identify frequent patterns.* Patterns included sequences of 2-5 actions. To ensure patterns considered for analysis were meaningful we only consider patterns that occur for at least 20% of students in a group. This threshold is the same as described in [9].
- *Calculate s-support and i-support metrics for each pattern.* Metrics were calculated for each group using the definition described above.
- *Identify statistically significant differences in frequency.* T-tests with a Bonferroni correction were conducted to compare the *s-support* and *i-support* metrics across each pair of SRL classifications. The Bonferroni correction was conducted for 95% confidence across the three pairwise tests but did not account for the multiple comparisons across patterns. This approach was employed because the primary purpose of our investigation was to identify meaningful patterns, not to prove statistical differences between populations [9].

5 Results

In total, 137 sequences were identified as frequent, occurring in more than 20% of student traces. Of these 29 were identified as having a significant difference in frequency between Low, Medium, or High SRL students. Further interpretation of these sequences suggested 6 general behavior patterns that occurred at different frequencies between the groups (Table 1). Of these, 3 patterns were more frequently displayed by High SRL students, while the remaining 3 patterns were more frequent among Low SRL students. These general patterns of behavior provide important insight into how students differentially interact with the environment given their level of SRL skill.

For instance, patterns **P1** and **P3** both highlight High SRL students' usage of the diagnosis worksheet. Specifically, these students are more likely to keep track of information as they receive it. Both the hypothesis and symptoms area of the diagnosis worksheet are optional, suggesting that High SRL students are choosing to use the resource to help themselves keep track of their ideas. Additionally, while the symptoms section of the worksheet involves simple recording of facts, the hypothesis area requires students to synthesize what they know and make inferences about the likelihood of different hypotheses, indicating strong inquiry skills. Together these patterns indicate that High SRL students are utilizing resources to keep track of what they know and are actively reflecting on the inquiry process.

In contrast, pattern **P5**, which is demonstrated more frequently by Low SRL students, indicates poor planning and inquiry skills. This pattern involves students reading about diseases, then visiting patients to ask about their symptoms, and repeating this process. This pattern suggests that Low SRL students are gathering data "just in time." They are repeatedly checking the information from patients against the information in books and posters to arrive at a hypothesis. These students are not keeping

Table 1. Differential patterns of behavior

	Sample Sequences	s-support			i-support		
		L	M	H	L	M	H
High SRL Students	**P1: Reading about diseases and updating hypotheses in worksheet**						
	READDIS-WSHYP-READDIS-WSHYP-READDIS	0.10	0.19	0.26	0.28	0.53	0.74
	WSHYP-READDIS-WSHYP-READDIS-WSHYP	0.11	0.18	0.24	0.31	0.54	0.70
	P2: Talk about problem and learn about pathogens						
	TALKPROB-TALKPATH-TALKPROB	0.75	0.80	0.88	0.83	0.95	0.98
	TALKPROB-TALKPATH-READPATH	0.15	0.25	0.27	0.16	0.26	0.27
	P3: Talk about symptoms and update symptoms in worksheet						
	TALKSYM-WSSYM	0.42	0.61	0.61	0.74	1.16	1.13
	TALKSYM-TALKPROB-WSSYM	0.03	0.08	0.12	0.03	0.08	0.13
Low SRL Students	**P4: Alternating incorrect and irrelevant tests**						
	TESTIRR-TESTINC-TESTIRR	0.61	0.55	0.47	1.79	1.55	1.01
	TESTINC-TESTIRR	0.71	0.71	0.66	2.27	2.04	1.50
	P5: Read about diseases and ask about symptoms						
	TALKSYM-READDIS-TALKSYM-READDIS	0.39	0.26	0.23	0.39	0.31	0.25
	READDIS-TALKSYM-READDIS-TALKSYM-TALKPROB	0.35	0.19	0.18	0.35	0.22	0.19
	P6: Learn about pathogens and run irrelevant tests						
	READPATH-TESTIRR-TESTINC	0.33	0.29	0.18	0.44	0.34	0.24
	TALKPROB-TALKPATH-TESTIRR	0.28	0.25	0.21	0.34	0.32	0.24

track of this information in their diagnosis worksheet and consequently are going back and forth between the books and posters on diseases to the infirmary with the patients. This represents a much less effective approach to problem solving when compared with the High SRL students. Additionally, these students are likely experiencing an increased cognitive load as they are trying to recall all the details they have gathered without the aid of the in-game resources. These patterns indicate that Low SRL students may need scaffolding for effective organization of knowledge and use of external cognitive tools, which is an important component of self-regulated learning [6, 10].

Another important distinction concerns students making connections about the type of illness affecting the patients. Specifically, students learn that the illness was spread through food that the camp members ate (TALKPROB). Students should also learn (through TALKPATH or READPATH) that a pathogen is a type of illness that can be spread through food or contact, whereas mutagens and carcinogens are not spread from person to person. Students should consequently conclude that the illness is a form of pathogen. This may be what is occurring in pattern **P2** demonstrated by High SRL students. These students are alternating between finding out information about the nature of the illness and about pathogens and are likely using this information to draw the conclusion that the illness is a form of pathogen. Additionally, the back-and-forth nature of these activities suggests goal-driven behavior perhaps to inform their testing strategies.

When running tests in the lab, students select whether they are testing for pathogens, mutagens or carcinogens. Knowledge of the pathogens and the nature of the illness should preclude students from running tests on carcinogens or mutagens (TEST$_{IRR}$); however, pattern **P6** indicates that Low SRL students are not making this connection or choose to ignore it. Additionally, **P4** suggests that Low SRL students may not be carefully selecting their testing strategy based on prior knowledge and may be trying any form of test to get a positive result. This suggests that Low SRL students may need more guidance in making the connection between the nature of the problem and type of illness. Additionally, they should be encouraged to identify whether the source is a pathogen, mutagen, or carcinogen before beginning to test.

6 Conclusion

Open-ended, inquiry-based learning environments are powerful tools for engaging students in scientific thinking and authentic problem solving. However, not all students are successfully able to navigate these environments and learn effectively. Self-regulated learning behaviors such as goal setting, progress monitoring, and effective tool use are critical for optimizing learning outcomes. Students lacing these skills have a disadvantage, but may be able to overcome this with additional guidance and support.

This work utilized differential sequence mining techniques to identify patterns of inquiry behaviors associated with self-regulated learning skills. Results indicated that students with more developed SRL skills utilize in-game resources more effectively to help reduce cognitive load. They also appear to be able to more effectively draw inferences and use them to inform future behaviors and strategies. These differences highlight areas for scaffolding students with less-developed regulatory skills. Specifically, Low SRL students can be encouraged and guided through the use of cognitive tools. Hopefully by clearly demonstrating how these tools can be successfully used, students will be more likely and more effective at using the resources. Additionally, it may be important to highlight ties between different sources of information and present specific learning goals related to each component of the problem solving activity. An important area for future work will be to incorporate these scaffolding strategies and to measure the impact on behavior patterns and overall learning for students who are not already strong self-regulated learners.

Acknowledgements. The authors wish to thank members of the IntelliMedia Group 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 Grants DRL-0822200, IIS-0812291, and CNS-0739216. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship.

References

1. Alfieri, L., Brooks, P., Aldrich, N. & Tenenbaum, H. Does Discovery-Based Instruction Enhance Learning. *Journal of Education Psychology* pp. 103, 1–18 (2011).
2. Kirschner, P. A., Sweller, J. & Clark, R. E. Why Minimal Guidance during instruction does not work: An analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist* 41, pp. 75–86 (2006).
3. Roll, I., Aleven, V. & Koedinger, K. R. The Invention Lab: Using a Hybrid of Model Tracing and Constraint-Based Modeling to Offer Intelligent Support in Inquiry Environments. *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* pp.115–124 (2010).
4. Woolf, B. P. *et al.* Critical Thinking Environments for Science Education. *Proceedings of the 12th International Conference on Artificial Intelligence in Education* pp. 515–522 (2005).
5. Ketelhut, D. J. The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in "River City", a multi-user virtual environment. *Journal of Science Education and Technology* 16, pp. 99–111 (2007).
6. Land, S. Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development* 48, pp. 61–78 (2000).
7. Schunk, D. H. Attributions as Motivators of Self-Regulated Learning. *Motivation and Self-Regulated Learning: Theory, Research, and Applications* pp. 245–266 (2008).
8. Ellis, D. & Zimmerman, B. J. Enhancing self-monitoring during self-regulated learning of speech. pp. 205–228 (2001).
9. Kinnebrew, J. S., Loretz, K. M. & Biswas, G. A Contextualized, Differential Sequence Mining Method to Derive Students' Learning Behavior Patterns. *Journal of Educational Data Mining* (in press).
10. Zimmerman, B. J. Self-regulated learning and academic achievement: An overview. *Educational Psychologist* 25, pp. 3–17 (1990).
11. Kostons, D., van Gog, T. & Paas, F. Training Self-Assessment and Task-Selection Skills: A Cognitive Approach to Improving Self-Regulated Learning. *Learning and Instruction* 22, pp. 121–132 (2012).
12. Cuevas, P., Lee, O., Hart, J. & Deaktor, R. Improving Science Inquiry with Elementary Students of Diverse Backgrounds. *Journal of Research in Science Teaching* 42, pp. 337–357 (2005).
13. Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C. & Greene, J. A. Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science* 33, pp. 381–412 (2005).
14. Biswas, G., Jeong, H., Roscoe, R. & Sulcer, B. Promoting Motivation and Self-Regulated Learning Skills through Social Interactions in Agent-Based Learning Environments. *2009 AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems* (2009).
15. Sabourin, J., Shores, L., Mott, B. & Lester, J. Predicting Student Self-Regulation Strategies in Game-Based Learning Environments. *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* pp. 470–475 (2012).
16. Rowe, J P, Shores, L. R., Mott, B. W., & Lester, J. C. Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. *International Journal of Artificial Intelligence in Education.* pp. 166–177 (2011).