Int J Artif Intell Educ (2013) 23:94–114
DOI 10.1007/s40593-013-0004-6

RESEARCH ARTICLE

Understanding and Predicting Student Self-Regulated Learning Strategies in Game-Based Learning Environments

Jennifer L. Sabourin · Lucy R. Shores · Bradford W. Mott · James C. Lester

Published online: 22 October 2013

Abstract Self-regulated learning behaviors such as goal setting and monitoring have been found to be crucial to students' success in computer-based learning environments. Consequently, understanding students' self-regulated learning behavior has been the subject of increasing attention. Unfortunately, monitoring these behaviors in real-time has proven challenging. This paper presents an initial investigation into self-regulated learning in a game-based learning environment. Evidence of goal setting and monitoring behaviors is examined through students' text-based responses to update their 'status' in an in-game social network. Students are then classified into SRL-use categories. This article describes the methodology used to classify students and discusses analyses demonstrating the learning and gameplay behaviors across students in different SRL-use categories. Finally, machine learning models capable of predicting these classes early in students' interaction are presented.

Keywords Self-regulated learning · Metacognition · Machine learning

Introduction

Understanding and facilitating students' self-regulated learning behaviors has been the subject of increasing attention in recent years. This line of investigation is fueled by evidence suggesting the strong role that self-regulatory behaviors play in a

J. L. Sabourin (✉) · L. R. Shores · B. W. Mott · J. C. Lester
North Carolina State University, Raleigh, NC, USA
e-mail: jlrobiso@ncsu.edu

L. R. Shores
e-mail: lrshores@ncsu.edu

B. W. Mott
e-mail: bwmott@ncsu.edu

J. C. Lester
e-mail: lester@ncsu.edu

student's overall (Zimmerman 1990). Self-regulated learning (SRL) can be described as "the process by which students activate and sustain cognitions, behaviors, and affects that are systematically directed toward the attainment of goals" and generally includes strategic, metacognitive, and motivational components (Schunk 2008 pg 245; Winne and Hadwin 1998). Unfortunately, students can demonstrate a wide range of fluency in their SRL behaviors (Ellis and Zimmerman 2001) with some students lagging behind their peers in their ability to appropriately set and monitor learning goals.

For this reason, the ability to identify and support students' SRL strategies has been the focus of much work in the AI in Education community (Azevedo et al. 2010; Biswas et al. 2009; Conati and VanLehn 2000). Such work has focused primarily on examining SRL in highly structured problem-solving and learning environments. However, understanding and scaffolding students' SRL behaviors is especially important in open-ended learning environments where goals may be less clear and students do not necessarily have a clear indicator of their progress. In order to be successful in this type of learning environment, students must actively identify and select their own goals and evaluate their progress accordingly. Unfortunately, students do not consistently demonstrate sufficient self-regulatory behaviors during interactions with these environments, which may reduce the potential contributions to learning (Alfieri et al. 2011; Kirschner et al. 2006). Consequently, further investigation of the role of SRL in open-ended learning environments is called for to understand how these environments can be used as effective learning tools.

This work describes a preliminary investigation of self-regulatory, and more specifically metacognitive, behaviors of students in a game-based science mystery, Crystal Island. During interactions with the Crystal Island environment, students were prompted to reflect on their mood and status in a way that is similar to many social networking tools available today. Though students were not explicitly asked about their goals or progress, many students included this information in their short, typed status statements. This data is used to classify students into low, medium, and high self-regulated learning behavior classes. Based on these classifications we investigate differences in student learning and in-game behaviors in order to identify the role of SRL in Crystal Island. Machine learning models are then trained that are capable of accurately predicting students' SRL-use categories early into their interaction with the environment, offering the possibility for timely intervention. The implications of these results and areas of future work are then discussed.

Background

Self-Regulated Learning

Self-regulated learning (SRL) is a term used to describe the behaviors of students who actively control their own learning (Schunk and Zimmerman 2003). More specifically, models of SRL are generally comprised of strategic, metacognitive, and motivational components manifesting as one's ability to evaluate and effectively control cognitive and motivational processes during learning within a particular domain (Zimmerman 2000; Pintrich 2000; Winne 2001). For example, self-regulated learners commonly set

their own learning goals and efficiently work toward them by carefully monitoring their progress and implementing adaptive strategies accordingly (Zimmerman 2000). Therefore, self-regulated learners are not only equipped with a sufficient set of learning strategies, but also have the motivational control to put forth the necessary effort to engage in these cognitive processes (Pintrich 2000). Consequently, there is evidence that students who are better able to regulate their learning in an intentional and reflective way often demonstrate greater academic motivation and achievement (Zimmerman 1990).

Social cognitive perspectives of learning posit SRL-related skills develop over time and are therefore malleable (Zimmerman 2002). However, while it seems most students perform self-regulatory behaviors during learning to some extent, the degree of competency is unfortunately broad, even among students of the same age (Ellis and Zimmerman 2001). To mediate these differences, Zimmerman (2002) suggests that self-regulatory behaviors should be modelled and encouraged by utilizing instructional methods that require choice and autonomy, allowing students to apply and reflect on related skills. This is motivated by evidence that the development of SRL-related skills is largely social and includes strategy use, metacognitive processes, and methods for maintaining motivation. For example, prompting students to explicitly explain their actions and thoughts has been shown to be an effective method for increasing students' metacognition and performance (Chi et al. 1989). Empirically, intervention research focused on process goals, feedback, and self-reflection in the classroom has yielded positive results (Schunk and Swartz 1993a, b; Schunk and Zimmerman 2003).

SRL and Intelligent Tutoring Systems

The affordances provided by intelligent tutoring systems provide an ideal platform for investigating and encouraging self-regulatory behaviors, and has thus been a focus of much work in this community. For example, in MetaTutor, a hypermedia environment for learning biology, think-aloud protocols have been used to examine which strategies students use, while analysis of students' navigation through the hypermedia environment helps to identify profiles of self-regulated learners (Azevedo et al. 2010). Further studies have examined the role of prompts encouraging students to set and monitor specific learning goals (Azevedo et al. 2012). The results of these investigations identified the importance of providing timely prompts as well as offering feedback on students' responses to the prompt.

Aleven and colleagues have examined modeling and scaffolding the self-regulatory process of help-seeking, self-explanation, and self-assessment in the Cognitive Tutor systems (Aleven et al. 2006). In these environments, students may request help at any time though many students do not use this feature correctly. Some students may request help too early or frequently without spending time on the problem, while others may avoid help even when it would be beneficial. Aleven et al. developed a model of ideal help seeking behavior to identify students who were not using the feature correctly. Using this model, the system delivered feedback either encouraging students to use or avoid help. While there were no learning differences between students who received feedback and those who did not, the authors did find that the feedback helped improve help-seeking behaviors in future interactions with the system. Furthermore, based on observations of students' problem-solving behaviors in the classroom (Chi et al. 1989),

integrating prompts for self-explanation within the Cognitive Tutor system have proven an effective method for increasing metacognition and, in turn, performance (Aleven and Koedinger 2002). Lastly, recent work by Long and Aleven (2013) found asking students to self-assess their problem solving skills yielded productive self-reflection, which positively affected learning outcomes and self-assessment accuracy.

Similarly, researchers have identified patterns of behavior in the Betty's Brain system that are indicative of low and high levels of self-regulation (Biswas et al. 2009). Prompting students to use SRL strategies when these patterns of behavior occur has shown promise in improving student learning. Additionally, Conati et al. have examined the benefits of prompting students to self-explain when learning physics content in a computer-based learning environment (Conati and VanLehn 2000).

While previous work has focused primarily on examining SRL in highly structured settings, open-ended exploratory environments provide choice and autonomy allowing students to engage in and practice SRL-related competencies. However, given the variance in the development of self-regulatory skills among students, the instructional efficacy of open-ended learning is highly dependent upon the environment's ability to encourage and scaffold metacognitive processes, effective strategy use, and motivation. For example, while the nature of the learning task may have implicit overarching goals such as 'completing the task' or 'learning a lot,' it is important for students to set more specific, concrete and measurable goals (Land 2000; Zimmerman 2008). As a result, such instructional interventions have garnered considerable attention within the SRL and intelligent tutoring system communities.

Game-based Learning

Game-based learning has been proposed as an approach to encourage positive affect, engagement, and motivation in learning activities by utilizing game-like features and environments (Gee 2003; Kapp 2012; Shaffer 2006). This work draws on empirical evidence that games are highly motivating and have natural ties with how people learn (Gee 2003; Kapp 2012; McNamara et al. 2009). In recent years, games have been used to teach a variety of subjects including scientific inquiry (Clarke and Dede 2009; Rowe et al. 2011), mathematics principles (Conati 2002), negotiation skills (Kim et al. 2009), foreign languages (Hallinen et al. 2009; Johnson 2010), policy argumentation (Easterday et al. 2011), and critical reasoning (Millis et al. 2011).

While some game-based learning systems have been shown to be effective in increasing student knowledge and fostering engagement and positive affect (Hallinen et al. 2009; Rowe et al. 2010), there are significant open questions about the efficacy of game-based learning (Mayer and Johnson 2010). Commonly, the competing findings documented in the literature can be explained through the environment's lack of support for students' self-regulation causing some students to flounder in the open-ended learning activities (Kirschner et al. 2006; Mayer and Johnson 2010; Easterday et al. 2011). For example, game features that are superfluous to the learning task designed to encourage interest and motivation may also introduce many distractions or "seductive details" that draw student attention away from the learning tasks (Harp and Mayer 1998; Sabourin et al. 2011b). According to Fiorella and Mayer (2012), "one of the challenges associated with the design of educational games is that the features intended to motivate students may result in extraneous processing" (p. 1076). In other

words, students may become distracted by the characters and objects that are present in the world or may spend time playing with aspects of the physics engine that underlies the gameplay. There is evidence that while these features may offer positive benefits for engagement, not all students are capable of regulating their behavior in a way that takes advantage of these features without harming their learning outcomes (Sabourin et al. 2011b).

Therefore, many have advocated the need to establish "the optimal balance between entertainment and education" for all learners (Mayer and Johnson 2010, p. 248), and recent work has begun to investigate intelligent SRL scaffolding and adaptation. One promising area of research involves leveraging *cognitive tools*—resources or processes seamlessly embedded within a learning environment designed to support students' cognitive processes—as a means for scaffolding student behaviours without foregoing entertainment (Lajoie and Derry 1993). For example, work by Shores and colleagues found that students who took advantage of cognitive tools embedded within a game-based learning environment reported higher levels of interest and demonstrated greater learning gains than students choosing to forego this scaffolding (Shores and Nietfeld 2011). Similarly, others have successfully modelled student behaviors and tool use as a means for assessing skills and performance stealthily in real-time without directly consulting the student (Baker and Clarke-Midura 2013; Shute and Ventura 2013). Consequently, investigations into the early prediction of students' cognitive tool use provide a potential avenue for promoting knowledge and skill acquisition through strategy use without compromising engagement (Shores et al. 2011). There is further evidence that students who utilize effective problem-solving strategies are more efficient problem solvers and report more positive affective outcomes as a result (Sabourin et al. 2012a, b).

This line of investigation has yielded promising preliminary results and calls for additional research into how best to recognize and support all three components of SRL while interacting with game-based learning environments. Therefore, from a metacognitive perspective, an important area of research includes extending findings in support of self-reflection for game-based learning (Mayer and Johnson 2010) to intelligent learning environments. The present investigation seeks to answer the following questions: 1) Can metacognitive status statements elicited through prompts embedded within the game narrative play provide meaningful measures of SRL? 2) How are these metacognitive measures related to learning, in-game behaviors, and motivation? and 3) Can machine learning techniques be used to predict these measures of SRL early into a student's interaction?

Crystal Island

An investigation of students' SRL behaviors was conducted with CRYSTAL ISLAND (Fig. 1), a game-based learning environment being developed for the domain of microbiology. CRYSTAL ISLAND features a science mystery designed to support the North Carolina eighth-grade microbiology curriculum. The premise of the mystery is that the student arrives on an island to discover that the research team that has been established there has fallen ill. The camp nurse explains that they have not been able to identify the cause or type of illness and asks for the student's help. The student then

Fig. 1 CRYSTAL ISLAND learning environment

works to collect clues by talking with virtual characters, running tests on objects in the world, and reading related books and posters. Once the student arrives at the correct source and type of illness and proposes a diagnosis, they have solved the mystery and completed the game.

There are a variety of activities students engage in to learn the material and solve the mystery. Students are encouraged to read books and posters with the related microbiology content. They must converse with characters to understand the symptoms of the ill camp members and to acquire information about what types of foods the camp members have been eating. They must gather and run tests on relevant items to determine what objects might be contaminated. Once a contaminated object has been found, they may examine it further with a microscope to distinguish with what it is contaminated. They are also encouraged to keep track of all their findings and hypothesis by taking notes and recording data in their *diagnosis worksheet*. CRYSTAL ISLAND offers a rich interactive world built using the Valve Software Source engine.

Method

Data was collected from 296 middle school students interacting with the CRYSTAL ISLAND environment. During their interactions, students were prompted to report on their mood and status in a way that is similar to many social networking tools available today. Though students were not explicitly asked about their goals or progress, many students included this information in their short, typed status statements. This data is used to classify students into low, medium, and high self-regulated learning behavior classes that are used for further analyses.

Participants

After removing instances of incomplete data, the final corpus included data from 260 students. Of these, there were 129 male and 131 female participants. The average age of the students was 13.4 years (SD=0.57). Approximately 1 % of the participants were American Indian or Alaska Native, 1 % were Asian, 17 % were Black or African American, 8 % were Hispanic or Latino, 63 % were Caucasian, and 10 % were of mixed or other races. At the time of the study, the students had not yet completed the microbiology curriculum in their classes.

Measures

A week prior to the interaction, students completed a set of pre-study questionnaires including a test of prior knowledge, as well as several measures of personal attributes. *Personality* was measured using the Big 5 Personality Questionnaire, which describes personality along five dimensions: openness, conscientiousness, extraversion, agreeableness and neuroticism (McCrae and Costa 1993). *Goal orientation*, which refers to the extent that a student values mastery of material and successful performance outcomes when engaged in learning activities, was also measured (Elliot and McGregor 2001). Finally, students' *emotion regulation* strategies were measured with the Cognitive Emotion Regulation Questionnaire (Gernefski and Kraati 2006), which measures the extent to which each of nine common strategies are used by an individual student. Students also completed a researcher-generated curriculum test to measure their level of knowledge before interacting with CRYSTAL ISLAND.

Immediately after completing their interaction with CRYSTAL ISLAND, students were given a post-interaction curriculum test with questions identical to that of the pre-test. Students also completed two questionnaires aimed to measure students' interest and involvement with CRYSTAL ISLAND. A shortened form of the Intrinsic Motivation Inventory (McAuley et al. 1989) was used to measure student motivation along five factors: interest/enjoyment, perceived competence, effort/importance, pressure/tension, and value/usefulness. Student engagement and presence in the system was measured using the Presence Questionnaire (Witmer, and Singer 1998), which includes several subscales such as a sense of immersion and involvement. Students also completed a questionnaire related to their understanding of the CRYSTAL ISLAND mystery, though these measures are outside the scope of this discussion.

Procedure

Having completed the pre-test materials a week prior to the study, the students were seated at a station that included a laptop, mouse, headphones and a set of explanatory materials. Students received an introductory presentation by a researcher, which included a brief description of the purpose of the study and details about the game controls. The explanatory materials at each student station included this same information printed so they could reference it throughout the study.

During the study, students interacted with CRYSTAL ISLAND for 55 min or until they completed the mystery. Student's in-game behaviors were logged in detail by the system. During their interaction, they also received an in-game prompt asking them to report on their emotional state (Fig. 2). This prompt was described to students as being part of an "experimental social network" that was being used on CRYSTAL ISLAND. They received the prompt every 7 min and could only make updates when prompted. Students selected from one of seven emotional states: anxious, bored, confused, curious, excited, focused, and frustrated. They were also asked to type a short "status update." After their interactions, students were directed to a computer lab where the completed the post-interaction materials.

Fig. 2 Self-report device

SRL Annotation

During their interactions with CRYSTAL ISLAND, students were occasionally prompted to indicate their emotional state and to briefly type a few words about their current "status," similarly to how they might update their status in an online social network. Though these prompts did not directly ask students to report on goal-setting or monitoring, many student responses contained evidence of these metacognitive behaviors. Student status reports were tagged for SRL evidence use using the following four ranked classifications: (1) *specific reflection*, (2) *general reflection*, (3) *non-reflective* statements, or (4) *unrelated* (Table 1). This ranking was motivated by the observation that setting and reflecting upon goals is a hallmark of self-regulatory behaviors and that specific goals are more beneficial than those that are more general (Zimmerman 2008).

From the final corpus of 260 students, a total of 1836 statements were collected, resulting in an average of 7.2 statements per student. All statements were tagged by one member of the research team with a second member of the research team tagging a randomly selected subset (10 %) of the statements to assess the validity of the protocol. Inter-rater reliability was measured at κ=0.77, an acceptable level of agreement. General reflective statements were the most common (37.2 %), followed by unrelated (35.6 %), specific reflections (18.3 %) and finally non-reflective statements (9.0 %).

Table 1 SRL tagging scheme

SRL category	Description	Examples
Specific reflection	Student evaluates progress towards a specific goal or area of knowledge	"I am trying to find the food or drink that caused these people to get sick." "Well…the influenza is looking more and more right. I think I'll try testing for mutagens or pathogens – [I] ruled out carcinogens"
General reflection	Student evaluates progress or knowledge but without referencing a particular goal	"I think I'm getting it" "I don't know what to do"
Non-reflective	Student describes what they are doing or lists a fact without providing an evaluation	"testing food" "in the lab"
Unrelated	Any statement which did not fall into the above three categories is considered unrelated, including non-word or unidentifiable statements	"having fun" "arghhh!"

After tagging, students were given an overall SRL score based on the average score of their metacognitive statements, with unrelated statements contributing 0 points through to specific reflection which contributed 3 points per statement. The average was taken since students had different frequencies of reports based on when they solved the mystery. Because students could not enter statements at will we did not consider the count of statements as indicative of metacognition. The average SRL score was 1.40 (SD=0.84) with a minimum and maximum score of 0 and 3, respectively. An even tertiary split was then used to assign the students to a Low, Medium, and High SRL category. This even split occurred at the levels of 1.0 and 2.0 (Fig. 3). This split implies that High SRL students are making mostly specific reflections, while Low SRL students are primarily un-reflective. Three groups were chosen to cover a larger gradient of SRL behaviors and because tertiary splits have been effective in other machine learning applications (Sabourin et al. 2011a; Shores et al. 2011).

Implications of SRL Classifications

Recall that the investigation sought to investigate whether metacognitive status statements elicited through prompts embedded within the game narrative play provide *meaningful* measures of SRL. While it was found that in-game prompts yielded evidence of metacognitive processes, it is not clear whether these measures are meaningful in predicting further outcomes. It was thus important to address the components of the second question by investigating how the metacognitive SRL classifications related to additional components of SRL such as learning, in-game behaviors and motivation.

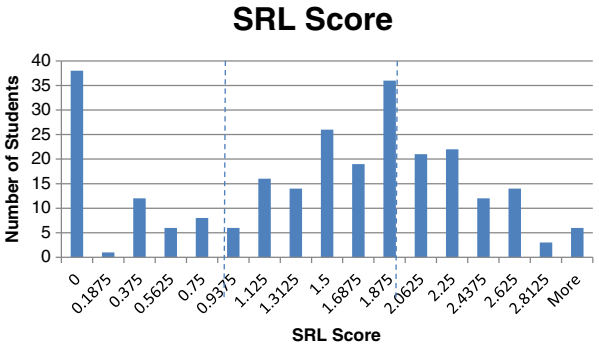

Fig. 3 Histogram of SRL scores

SRL and Learning

The first objective of this work was to understand the role self-regulatory tendencies played on student learning. First, pre-curriculum test scores and SRL scores were group-mean centered to establish a meaningful midpoint. Subsequently, a hierarchical linear regression predicting students' performance on the post-curriculum test was performed by entering pre-test scores into the first block and the SRL score into the second block. While pre-test scores were found to be significantly predictive of post-test scores ($F_{(1,\ 258)} = 51.07$, $p<.0001$, $r^2=.16$), students' SRL scores were also significantly predictive above and beyond prior knowledge ($F_{(2,\ 257)} = 39.63$, $p<.0001$, $r^2=.24$) accounting for approximately 8 % more of the variance. Furthermore, student learning, as measured by normalized learning gains from the pre-test to post-test, was compared for the three SRL groups. An ANOVA indicated a significant difference in learning gains between the groups ($F_{(2,\ 257)} = 4.6$, $p<0.01$). Tukey post-hoc comparisons indicated that both High and Medium SRL students experienced significantly better learning gains than Low SRL students at the $\alpha=0.05$ level. Analyses also indicated that there were significant differences on pre-test scores between groups ($F_{(2,\ 257)} = 5.07$, $p<0.01$) suggesting that students with high SRL tendencies may be better students or perhaps their increased prior knowledge helped them to identify and evaluate their goals more efficiently. Figure 4 shows the pre- and post-test scores across groups, highlighting both the differences in prior knowledge and learning during interaction with CRYSTAL ISLAND.

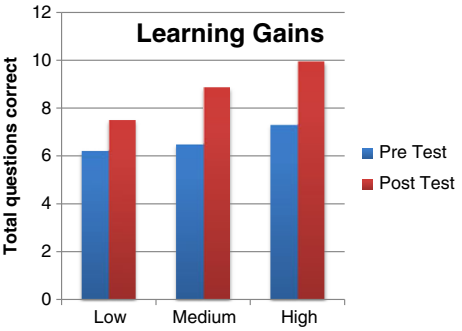

Fig. 4 Pre- and post-test scores by SRL classification

SRL and In-Game Behaviors

The next set of analyses was conducted to investigate differences in student behavior based on their SRL tendencies. A chi-squared analysis indicated that the percentage of students who solved the mystery did not differ significantly based on SRL group (χ^2 (2, N=260) = 4.72, p=0.094) though on average, 38 % of High, 38 % of Medium and 25 % of Low SRL students completed the mystery. Additionally, an ANOVA found there was no significant difference in the number of goals completed during the interaction.

While a significant difference in students' abilities to solve the mystery was not found, there were differences in the in-game resources that students used. Resources expected to be most beneficial to learning and self-regulation included a *microbiology app* on the students' in-game smartphone which provides a wealth of microbiology information, *books* and *posters* that are scattered around the island with additional information, a notebook where students can record their own *notes*, a structured *diagnosis worksheet* for recording findings and hypotheses and finally a *testing* machine where students formulate hypotheses and run relevant tests. ANOVAs for student use of each of these features indicated a significant difference in student use of *posters* ($F_{(2,\ 257)}$ = 5.28, p<0.01), and *tests* ($F_{(2,\ 257)}$ = 5.59, p<0.01). While the differences in the use of other devices were not significant, interesting trends emerged (Fig. 5). High SRL students appear to make more use of the curricular resources in the game such as books and posters and also take more notes than the lower SRL students. Interestingly, High SRL students run significantly fewer tests than Medium or Low SRL students (as indicated by Tukey post-hoc comparisons). Abundant use of the testing device is often indicative of students gaming the system or failing to form good hypotheses in advance. This finding suggests that High SRL students may be more carefully selecting which tests to run and are perhaps obtaining positive test results earlier than Medium and Low SRL students.

SRL and Motivation

A final set of analyses was conducted to determine the role of SRL and motivation with the CRYSTAL ISLAND environment. The first measure of motivation that was

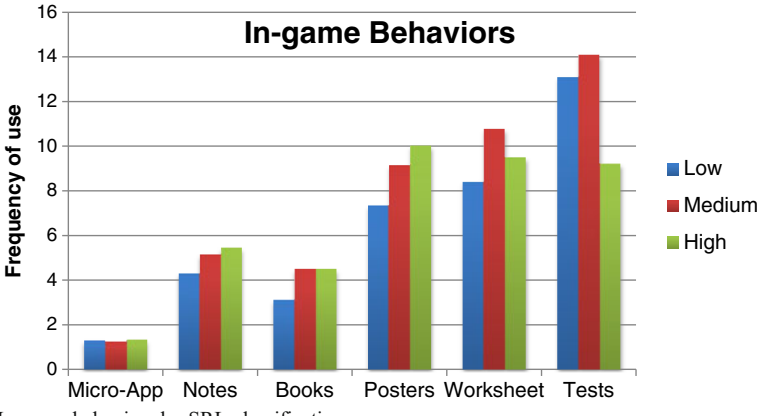

Fig. 5 In-game behaviors by SRL classification

considered was the total amount of off-task behavior students displayed during game-play as disengagement from the task is thought to be evident of low motivation and persistence (Hershkovitz et al. 2012). Off-task behavior is measured as instances where students disengage from the learning content and focus instead on the game-based features of the environment (Sabourin et al. 2011b). For example, activities such as stacking boxes, climbing on trees, and spending too much time at the waterfall were considered as off-task. Although there appeared to be a trend in which higher SRL students engaged in less off-task behavior, an ANOVA indicated that this difference was not significant, ($F_{(2,\ 257)} = 2.00$, p=0.13).

Further analyses were conducted with subscales of the post-measures of the Intrinsic Motivation Inventory (Fig. 6). ANOVA analyses indicated significant differences in the subscales of interest/enjoyment ($F_{(2,\ 257)} = 9.44$, p<0.01), effort/importance ($F_{(2,\ 257)} = 8.86$, p<0.01), and value/usefulness ($F_{(2,\ 257)} = 12.23$, p<0.01). Tukey post-hoc comparisons indicated that High SRL students reported experiencing significantly more interest and enjoyment, and attributed greater value and importance to the task than either Medium or Low SRL students. It is interesting to note that there was no significant difference in the perceived competence subscale ($F_{(2,\ 257)} = 2.07$, p<0.13). This was surprising due to the significant difference in learning gains demonstrated by students; however this finding may be explained by the lack of significant difference in the frequency of students completing the mystery within the given time. Perhaps students judged their competence on game skills specifically and not on the learning content. An alternate explanation could be that students with less developed SRL skills may be poor judges of their own competence (Kostons et al. 2012) and may over-estimate their success in this environment.

Finally, analyses were conducted on two subscales of the Presence Questionnaire believed to be most relevant to student motivation: involvement and interest. ANOVAs indicated significant differences in both of these metrics, ($F_{(2,\ 257)} = 4.59$, p=0.01) and ($F_{(2,\ 257)} = 6.44$, p<0.01), respectively. Tukey post-hoc comparisons indicated that High SRL students reported experiencing significantly more immersion and involvement than either Medium or Low SRL students.

These results highlight several important factors relating to the research questions guiding the investigation. First, the post-interaction method of classifying students

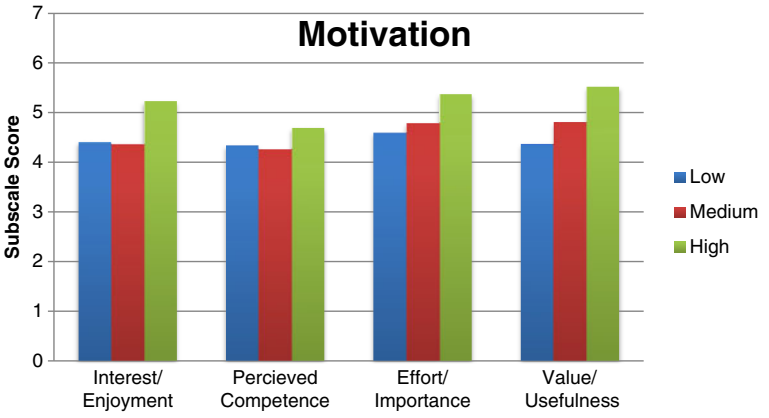

Fig. 6 Reported motivation by SRL classification

into Low, Medium, and High SRL categories appears to yield meaningful groupings of students. Second, these classifications have significant implications for student learning. Students in the High SRL group have a higher level of initial knowledge than Low SRL students and, through interactions with CRYSTAL ISLAND, increase this gap in knowledge. This highlights the importance of identifying the Low SRL students so they can be provided supplementary guidance to help bridge the gap. Additionally, the results indicate that High SRL students utilize the environment's curricular features differently and likely more effectively than Low SRL students. This finding suggests that scaffolding to direct Low SRL students towards more effective use of these resources could be an appropriate mechanism for bridging the learning gap. Finally, High SRL students also report increased levels of engagement and motivation over other students. Overall, these findings highlight the need to provide support that is tailored to the individual students' SRL abilities.

Predicting Self-Regulation Behaviors

In order to provide adaptive support of SRL strategies, students must be classified early into their interaction with the CRYSTAL ISLAND environment. The current procedure for identifying these students is performed manually after the interaction has been completed, which does not allow for early interventions. It is also desirable to only provide scaffolding specifically tailored to a student's skills since student learning and engagement may potentially be harmed by superfluous or inappropriate interventions. For these reasons, the next goal of this research was to train machine-learning models to predict students' SRL-use categories early into their interaction with CRYSTAL ISLAND. Because the primary goal of the learning environment is to improve science learning, appropriately recognizing and scaffolding the Low-SRL class of students is of increased importance.

Features

In order to predict students' SRL-use categories, a total of 49 features were used to train machine-learning models (Table 2). Of these, 26 features represented personal data collected prior to the students' interaction with CRYSTAL ISLAND. This included demographic information, pre-test score, and results of the personality, goal orientation, and emotion regulation questionnaires. The remaining 23 features represented a summary of students' interactions in the environment. This included information on how students used each of the curricular resources, how many in-game goals they had completed, as well as evidence of off-task behavior. Additionally, data from the students' self-reports were included, such as the most recent emotion report and the character count of their "status."

In order to examine early prediction of students' SRL-use categories, these features were calculated at four different points in time resulting in four distinct datasets. The first of these (**Initial**) represented information available at the beginning of the student's interaction and consequently only contained the 26 personal attributes. Each of the remaining three datasets ($\mathbf{Report_{1–3}}$) contained data representing the student's progress at each of the first three emotion self-report instances. These

Table 2 Features for trained models

Personal attributes (26)	In-game attributes (23)
Achievement Goals Questionnaire (4)	Resource Use (5)
• Mastery Avoidance	• Posters
• Master Approach	• Books
• Performance Avoidance	• Microbiology App
• Performance Approach	• Notes
Big Five Inventory (5)	• Tests Run
• Openness	Worksheet Use (6)
• Conscientiousness	• Worksheet Access
• Extraversion	• Worksheet Updates
• Agreeableness	• Worksheet Checks
• Neuroticism	• Worksheet Filled
Cognitive Emotion Regulation Questionnaire (9)	• Worksheet Right
• Self-Blame	• Worksheet Wrong
• Acceptance	Goal Completion (9)
• Planning	• Total Goals Completed
• Positive Refocusing	• Time Since Last Goal Completion
• Perspective Taking	• Talked to Kim
• Catastrophizing	• Talked to Teresa
• Other-Blame	• Talked to Quentin
• Ruminating	• Talked to Robert
Demographics (8)	• Talked to Ford
• Pre-Test	• Talked to Bryce
• Age	• Successful Test
• Gender	Other Behaviors (3)
• Gaming Frequency (numeric)	• Average Status Length
• Gaming Frequency (class)	• Total Off-Task Behavior
• Gaming Hours	• Recently Off-Task Behavior
• Gaming Skill (numeric)	
• Gaming Skill (class)	

datasets contained the same 26 personal attributes, but the values of the remaining 23 in-game attributes differentially reflected the student's progress up until that point. The first self-report occurred approximately 4 min into game play with the second and third reports occurring at 11 min and 18 min, respectively. The third report occurs after approximately one-third of the total time allotted for interaction has been completed, so it is still fairly early into the interaction time.

Predictive Modeling

Each of the described datasets was used to train a set of machine learning classifiers including Naïve Bayes, Decision Tree, Support Vector Machine, Logistic Regression, and Neural Network. These models were trained and evaluated using 10-fold cross-validation

Table 3 Predictive models and evaluation metrics (for predictive accuracy, * and ** indicate a significant improvement over the prior prediction at $p<.05$ and .01, respectively)

Model	Predictive accuracy				Low-SRL recall			
	Initial	$Report_1$	$Report_2$	$Report_3$	Initial	$Report_1$	$Report_2$	$Report_3$
Naïve Bayes	44.2	43.5	46.1*	50.5*	0.47	0.28	0.54	0.52
Neural network	42.3	43.8	46.5*	45.5	0.44	0.45	0.49	0.52
Log. reg.	42.7	51.2**	47.7	54.5**	0.45	0.65	0.66	0.73
SVM	43.5	46.9*	45.7	51.4**	0.51	0.55	0.56	0.62
Decision tree	42.7	46.2*	48.1*	57.2**	0.45	0.55	0.71	0.71

with the WEKA machine learning toolkit (Hall et al. 2009). The predictive accuracies of these models are shown in Table 3. All of the learned models were able to offer a predictive accuracy statistically significantly better than a most-frequent class baseline (at $p<0.01$). Due to the fact that the classes were identified using an even tertiary split, the most frequent class (Medium) model has a predictive accuracy of 33.5 %. Additionally, most models demonstrated gains in predictive accuracy further into the interaction (Fig. 7).

Of the models attempting to predict SRL class before any interaction with the environment, the model with the best performance is the Naïve Bayes model (44.3 %). However, there are no significant differences in predictive accuracy between any of the models trained on this dataset. Alternatively, of the models trained with the most data, the Decision Tree model achieves the highest predictive accuracy (57.2 %), and is statistically significantly better than the other models trained on this dataset ($p<0.05$). In general, it appears that the two models with the best overall performance are the Decision Tree and Logistic Regression models.

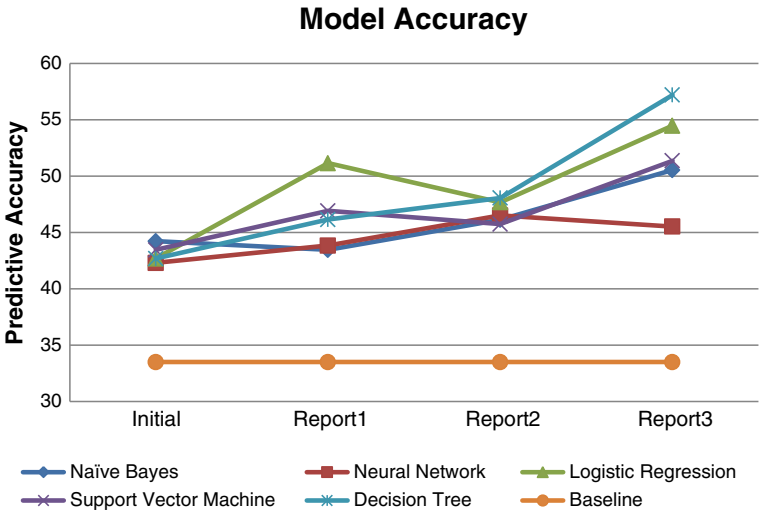

Fig. 7 Predictive accuracy improvements across time

In addition to predictive accuracy, we are also particularly interested in the models' abilities to distinguish Low SRL students as these students would be the targets of additional support. For this reason, we compared the models' levels of recall for the Low SRL class (Fig. 8). These results again demonstrate a steady growth in the ability to correctly recognize Low SRL students. Additionally, the Decision Tree and Logistic Regression models again distinguish themselves in their ability to outperform the remaining models. These results indicate that using either model, or perhaps a combination of both models, will offer promise in being able to identify and support Low SRL students early into their interaction with CRYSTAL ISLAND.

Discussion

This work presents an initial analysis of students' natural self-regulated learning activities in the game-based learning environment, CRYSTAL ISLAND. Three questions guided this investigation: 1) Can metacognitive status statements elicited through prompts embedded within the game narrative play provide meaningful measures of SRL?, 2) How are these metacognitive measures related to learning, in-game behaviors, and motivation?, and 3) Can machine learning techniques be used to predict these measures of SRL early into a student's interaction? Results indicate that undirected prompts have the potential to show students' use of goal setting and monitoring. Additionally, classifications using these metacognitive statements yielded meaningful groupings of students who demonstrated different levels of learning, in-game strategy use, and motivation. Specifically, the findings suggest that self-regulated learners tend to make better use of in-game curricular resources and may be more deliberate in their actions. Although highly self-regulated learners were not more likely to solve the mystery, they did demonstrate significantly higher learning gains as a result of their interaction. These results point to the importance of being able to identify students with tendencies towards low self-regulation in order to provide appropriate scaffolding.

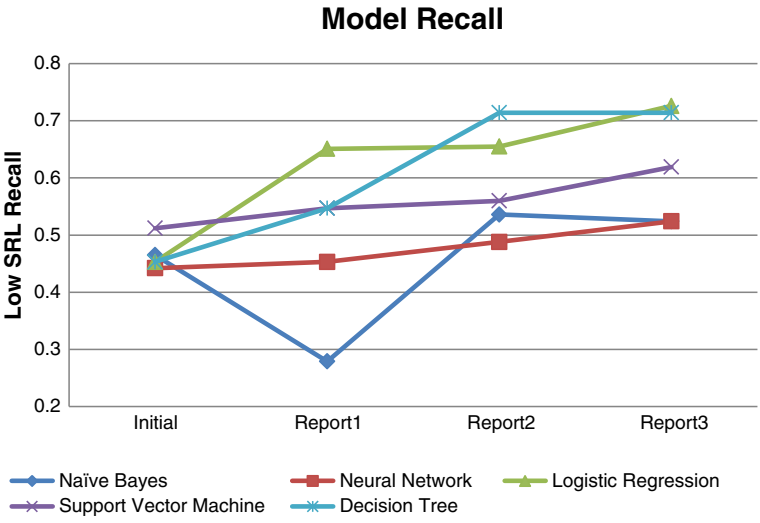

Fig. 8 Low-SRL recall improvements across time

Finally, the machine learning models discussed in this paper show significant promise in being able to predict a student's SRL abilities early into their interaction with CRYSTAL ISLAND.

Limitations

There are several limitations to this work. The primary limitation relates to the collection of evidence of self-regulated learning. The prompts were given to students in the context of a social network and did not directly prompt for reflection on goals or progress monitoring. We argue that students who made statements along these lines without prompting were likely actively engaging in goal setting and monitoring and had these reflections readily available. However, it is possible that we would see very different response patterns if the prompts had been guided and directed toward metacognitive monitoring. There may have been students who were engaging in these processes but did not choose to record this information in their status updates. However, it is also true that adding directed prompts will encourage students to engage in metacognitive processes that they were not predisposed to and may in turn impact the self-regulated learning behaviors that are exhibited. An important area of work will be to determine the role of directed and undirected prompts in identifying goal-setting and monitoring behavior.

Another limitation is the use of a tertiary split of students into High, Medium and Low categories. Because the modeling techniques used require categorical labels it was necessary to divide the students into some set of classes. The even three-way split was chosen because of prior work identifying this as providing a nice level of granularity without too many classes (Sabourin et al. 2011a); however, this selection does treat students in the upper and lower bounds of the bucket as equals while differentiating between students at the high and low endpoint of adjacent groups. It will be important to investigate other splitting techniques to identify optimal groupings of students. This may be done using clustering or other approaches that do not directly rely on a single measure.

Future Work

Many areas remain for future work. One major avenue for exploration is that of real-time recognition and scaffolding of SRL. Along these lines, it will be important to first identify which specific behaviors should be supported or guided by the adaptive system. This work has shown that High, Medium, and Low-SRL students utilize the features of the CRYSTAL ISLAND environment differently. Further work should be undertaken to attempt to gain a more detailed understanding of these differences with modeling techniques such as pattern mining or Markovian approaches. Next, leveled scaffolding will be developed and evaluated to identify how much scaffolding is appropriate for each SRL skill level. This scaffolding will encourage goal setting and monitoring behaviors and guide students towards strategies identified by the analysis of real student behaviors. It will be important to measure outcomes in terms of both learning and engagement as it is expected that too much guidance or support may reduce interest and enjoyment. Furthermore, it will be important to investigate the relative cost of misclassification and incorrect delivery of scaffolding. An objective cost

metric balancing engagement and learning can guide learned models towards policies that optimize a scaffolding strategy. Finally, the findings from each of these investigations will be incorporated into a comprehensive version of CRYSTAL ISLAND, capable of early detection and adaptive, leveled scaffolding of self-regulate learning.

Another interesting avenue is examining how these results generalize to other domains and environments. SRL skills such as goal setting and cognitive tool use have been shown to be important for success both in CRYSTAL ISLAND and in many other learning environments (Shores et al. 2011; Baker and Clarke-Midura 2013; Shute and Ventura 2013). It will be interesting to see how the features utilized for predictive modeling in this work can be extended to other environments with different interactive elements and structures. Identifying the commonalities between these environments for effective modeling will help in building more robust and generalizable models of self-regulated learning.

Conclusion

Self-regulated learning is an important skill impacting the success of students on a variety of learning tasks. Students without these skills are unable to make the most of learner-guided environments that provide autonomy and self-guided learning in the hopes of increasing engagement and interest as well as learning outcomes. Scaffolding tailored specifically to the skill-level of the student is necessary to balance the engagement benefits of autonomy and the learning benefits of guided learning activities. The empirical models discussed in this work represent the first step in developing a system capable of early identification of SRL skills so that adaptation can be tailored directly based on students' specific needs.

Acknowledgments 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 SourceTM engine and SDK. This research was supported by the National Science Foundation under Grants REC-0632450, DRL-0822200, and IIS-0812291. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. 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

Aleven, V., & Koedinger, K. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147–179.

Aleven, V., McLaren, B. M., Roll, I., & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: a model of help-seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101–128.

Alfieri, L., Brooks, P., Aldrich, N., & Tenenbaum, H. (2011). Does discovery-based instruction enhance learning. *Journal of Education Psychology, 103*(1), 1–18.

Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools. In M. Khine & I. Saleh (Eds.), *New science of learning: Cognition, computers and collaboration in education* (pp. 225–248). New York: Springer.

Azevedo, R., Landis, R., Feyzi-Behnagh, R., Duffy, M., Trevors, G., Harley, J., et al. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning

with MetaTutor. *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 212–221).

Baker, S. J. D. R., & Clarke-Midura, J. (2013). Predicting successful inquiry learning in a virtual performance assessment for science. In *Proceedings of the 21st International Conference on User Modelling, Adaptation, and Personalization* (pp. 203–214).

Biswas, G., Jeong, H., Roscoe, R., & Sulcer, B. (2009). 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*.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how student study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.

Clarke, J., & Dede, C. (2009). Design for scalability: a case study of the river city curriculum. *Journal of Science Education and Technology, 18*(4), 353–365.

Conati, C. (2002). Probabilistic assessment of user's emotions in educatioanl games. *International Journal of Applied Artificial Intelligence, Special Issue on Merging Cognition and Affect in HCI, 16*(1), 555–575.

Conati, C., & VanLehn, K. (2000). Towards computer-based support of meta-cognitive skills: a computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11*, 398–415.

Easterday, M. W., Aleven, V., Scheines, R., & Carver, S. M. (2011). *Using tutors to improve educational games. Proceedings of the 15th International Conference on Artificial Intelligence in Education* (pp. 63–71). Berlin: Springer Verlag.

Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501–519.

Ellis, D., & Zimmerman, B. J. (2001). In H. J. Hartman (Ed.), *Enhancing self-monitoring during self-regulated learning of speech* (pp. 205–228). Dordrecht: Kluwer.

Fiorella, L., & Mayer, R. E. (2012). Paper-based aids for learning with a computer-based game. *Journal of Educational Psychology, 104*(4), 1074–1082. doi:10.1037/a0028088.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy.* New York: Palgrave Macmillan.

Gernefski, N., & Kraati, V. (2006). Cognitive emotion regulation questionnaire: development of a short 18-item version. *Personality and Individual Differences, 41*, 1045–1053.

Hall, M., Frank, E., Holmes, G., Pfajhringer, B., Reutmann, P., & Witten, I. (2009). The WEKA data mining software: an update. *SIGKKD Explorations*, *11*(1).

Hallinen, N., Walker, E., Wylie, R., Ogan, A., & Jones, C. (2009). I was playing when I learned: a narrative game for French aspectual distinctions. *Proceedings of the Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education* (pp. 117–120).

Harp, S., & Mayer, R. E. (1998). How seductive details do their damage: a theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414–434.

Hershkovitz, A., Baker, R. S. J. D., Gobert, J., & Nakama, A. (2012). A data-driven path model of student attributes, affect, and engagement in a computer-based science inquiry microworld. *Proceedings of the International Conference on the Learning Sciences*.

Johnson, W. L. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education, 20*(2), 175–195.

Kapp, K. (2012). *The gamification of learning and instruction*. San Francisco: Pfeiffer.

Kim, J., Hill, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., et al. (2009). BiLAT: a game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education, 19*(3), 289–308.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). 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*, 75–86.

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: a cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*, 121–132.

Lajoie, S., & Derry, S. J. (1993). *Computers as cognitive tools*. Hilldale: Lawrence Erlbaum Associates.

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

Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In *Proceedings of the International Conference on Artifical Intelligence in Education (AIED-2013)* (pp. 219–228).

McAuley, E., Duncan, T., & Tammen, V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factory analysis. *Research Quarterly for Exercise and Sport, 60*(1), 48–58.

McCrae, R., & Costa, P. (1993). *Personality in adulthood: A five-factor theory perspective* (2nd ed.). New York: Guilford Press.

McNamara, D. S., Jackson, G. T., & Graesser, A. C. (2009). Intelligent tutoring and games (ITaG). *Proceedings of the Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Educationntelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education* (pp. 1–10).

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. (2011). Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious games and edutainment applications* (pp. 169–195). London: Springer.

Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*. San Diego: Academic.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). Integrating learning and engagement in narrative-centered learning environments. *Proceedings of the Tenth International Conference on Intelligent Tutoring Systems* (pp. 166–177). Pittsburgh, PA. Retrieved from http://www.springerlink.com/index/XJ5041R67060V8Q1.pdf.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education, 21*, 115–133.

Sabourin, J. L., Mott, B. W., & Lester, J. C. (2011a). Modeling learner affect with theoretically grounded dynamic Bayesian networks*, Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction*.

Sabourin, J. L., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011b). When off-task in on-task: the affective role of off-task behavior in narrative-centered learning environments. *Proceedings of the 15th International Conference on Artificial Intelligence and Education* (pp. 534–536). Springer-Verlag.

Sabourin, J., Rowe, J. P., Mott, B. W., & Lester, J. C. (2012a). Exploring inquiry-based problem-solving strategies in game-based learning environments. *Proceedings of the Eleventh International Conference on Intelligent Tutoring Systems* (pp. 470–475).

Sabourin, J., Rowe, J. P., Mott, B. W., & Lester, J. C. (2012b). Exploring affect and inquiry in open-ended game-based learning environments. *In Workshop on Emotions in Games for Learning in conjunction with the 11th International Conference on Intelligent Tutoring Systems*.

Schunk, D. H. (2008). Attributions as motivators of self-regulated learning. In Dale H. Schunk & B. J. Zimmerman (Eds.), *Motivation and Self-Regulated Learning: Theory, Research, and Applications* (pp. 245–266).

Schunk, D. H., & Swartz, C. W. (1993a). Goals and progress feedback: effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18*(3), 337–354.

Schunk, D. H., & Swartz, C. W. (1993b). Writing strategy instruction with gifted students: effects of goals and feedback on self-efficacy and skills. *Roeper Review, 15*(4), 225–230.

Schunk, D. H., & Zimmerman, B. J. (2003). In W. M. Reynolds & G. E. Miller (Eds.), *Self-regulation and learning* (Vol. 7, pp. 59–78). New York: Wiley & Sons.

Shaffer, D. W. (2006). *How computer games help children learn*. New York: Palgrave Macmillan.

Shores, L. R., & Nietfeld, J. L. (2011). The role of compensatory scaffolds for inquiry learning in narrative-centered learning environments. The 2011 Annual Meeting of the American Educational Research Association (AERA-2011). New Orleans, LA.

Shores, L. R., Rowe, J. P., & Lester, J. C. (2011). *Early prediction of cognitive tool use in narrative-centered learning environments. Proceedings of the Fifteenth International Conference on Artificial Intelligence in Education*. New Zealand: Auckland.

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment*. Cambridge: The MIT Press.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 153–189). Mahwah: Lawrence Erlbaum Associates.

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Mahway, NJ: Erlbaum.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: an overview. *Educational Psychologist, 25*, 3–17.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaert, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulated learning* (pp. 13–39). San Diego: Academic.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: an overview. *Theory into Practice, 41*(2), 64–70.

Zimmerman, B. J. (2008). Goal setting: A key proactive source of academic self-regulation. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 267–286). New York: Routledge.