ARTICLE IN PRESS

Learning and Instruction xxx (2017) 1–11

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment

Michelle Taub [a,*], Roger Azevedo [a], Amanda E. Bradbury [a], Garrett C. Millar [a], James Lester [b]

[a] Department of Psychology, North Carolina State University, 2310 Stinson Drive, Raleigh, NC 27695-7650, United States
[b] Department of Computer Science, North Carolina State University, 890 Oval Drive, Raleigh, NC 27695-8206, United States

A R T I C L E I N F O

Article history:
Received 28 December 2016
Received in revised form
12 June 2017
Accepted 23 August 2017
Available online xxx

Keywords:
Metacognition
Self-regulated learning
Scientific reasoning
Game-based learning
Sequence mining
Process data
Log files

A B S T R A C T

The goal of this study was to assess how metacognitive monitoring and scientific reasoning impacted the efficiency of game completion during learning with Crystal Island, a game-based learning environment that fosters self-regulated learning and scientific reasoning by having participants solve the mystery of what illness impacted inhabitants of the island. We conducted sequential pattern mining and differential sequence mining on 64 undergraduate participants' hypothesis testing behavior. Patterns were coded based on the relevancy of what items were being tested for, and the items themselves. Results revealed that participants who were more efficient at solving the mystery tested significantly fewer partially-relevant and irrelevant items than less efficient participants. Additionally, more efficient participants had fewer sequences of testing items overall, and significantly lower instance support values of the PartiallyRelevant–Relevant to Relevant–Relevant and PartiallyRelevant–PartiallyRelevant to Relevant–Partially Relevant sequences compared to less efficient participants. These findings have implications for designing adaptive GBLEs that scaffold participants based on in-game behaviors.

1. Introduction

Self-regulated learning (SRL) is an effective way of learning for students of all ages (Azevedo, 2014; Winne & Azevedo, 2014). When students self regulate their learning, they are playing an active role in the learning process by engaging in different cognitive, affective, metacognitive, and motivational (CAMM) processes (Azevedo, Taub, & Mudrick, 2015). Research has shown that using different self-regulatory skills can enhance learning (Winne & Azevedo, 2014), however upon investigating how students use these skills in the classroom, research has revealed that students are often unsuccessful at self regulating their learning effectively and efficiently (Azevedo et al., 2015; Lester, Rowe, & Mott, 2013). As such, researchers have developed different types of advanced learning technologies (ALTs) designed to foster effective SRL (Azevedo et al., 2015; Biswas, Segedy, & Bunchongchit, 2016; Graesser, 2013; Lester et al., 2013).

One specific category of ALTs is game-based learning environments (GBLEs), which were designed to foster engagement and enjoyment during gameplay and learning (e.g., Crystal Island, Alien Rescue, Cache 17, iSTART). For this study, we investigated how participants used SRL and scientific reasoning processes (i.e., hypothesis testing) during learning with Crystal Island. We assessed hypothesis testing within the game, which we contextualized as testing food items for the possible transmission source of the mysterious illness that impacted the inhabitants of the island. In addition, we made the assumption that what was occurring between testing events involved SRL, specifically metacognitive monitoring processes and knowledge acquisition. This was contextualized within the game as time spent reading virtual books and posters for knowledge acquisition, time spent talking to non-player characters (NPCs) that could help further narrow down the transmission source, and frequency of tracking and monitoring food items being tested into a diagnosis worksheet. A major concern when assessing learning with GBLEs for scientific reasoning is the issue of efficiency in terms of participants choosing the relevant evidence, making appropriate inferences and hypotheses, and testing relevant evidence, while still enjoying the game. SRL researchers have not addressed this concern, nor have

* Corresponding author.
E-mail addresses: mtaub@ncsu.edu (M. Taub), razeved@ncsu.edu (R. Azevedo), aebradbu@ncsu.edu (A.E. Bradbury), gcmillar@ncsu.edu (G.C. Millar), lester@ncsu.edu (J. Lester).

http://dx.doi.org/10.1016/j.learninstruc.2017.08.005
0959-4752/

Please cite this article in press as: Taub, M., et al., Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment, *Learning and Instruction* (2017), http://dx.doi.org/10.1016/j.learninstruc.2017.08.005

they investigated efficiency during gameplay. Therefore, for our analyses, we investigated the efficiency of scientific reasoning via hypothesis testing, and potential influences of SRL strategies (i.e., knowledge acquisition and monitoring processes) on the efficiency of hypothesis testing.

1.1. Theoretical framework

As our theoretical framework, we use the information processing theory (IPT) of self-regulated learning (Winne & Hadwin, 1998, 2008), which states that learning occurs through four cyclical stages: definition of the task, setting goals and plans, using learning strategies, and making adaptations to those goals, plans, and strategies, and, information processing, via the use of cognitive, metacognitive, and motivational SRL strategies, occurs during each stage. We use this model because each of these phases can relate to our study. In the first phase (definition of the task), students ensure that they are aware of what the task is asking them to do (e.g., solve the mystery of what disease has impacted all inhabitants by gathering clues from testing for the transmission source). In phase 2 (setting goals and plans), students set goals for how they plan to accomplish the task, as well as their plans for achieving those goals. For example, a student can set goals for gathering clues to help them solve the mystery, and their plans to do so would be to test different food items for the disease's transmission source, reading books to read about different diseases, including the symptoms associated with them, and talking to non-player characters who are sick patients and can report their symptoms, allowing students to match these symptoms to the ones they read about in the books, experts in microbiology who can tell them more about microbiology so they can narrow down if the disease is viral or bacterial, and workers on the island (i.e., camp cook) who could give more information about what food items inhabitants had been eating. The third phase, using learning strategies, involves students using cognitive and metacognitive strategies to enact the plans they set in the second phase. For this study, students could engage in cognitive learning strategies by reading information from the books and having conversations with non-player characters for knowledge acquisition about microbiology, and use metacognitive monitoring strategies to monitor the food items they were testing and their likelihood of being the disease's transmission source. It is important to note that knowledge acquisition is not an SRL strategy, however if students self-regulate their learning during knowledge acquisition, this can be beneficial because it can enhance knowledge acquisition by allowing students to actively acquire the information they need to complete the task. The fourth phase, making adaptations, involves the students adapting their goals, plans, and use of cognitive and metacognitive strategies based on their progression through the task. For example, students could decide to test all food items they found on the island, but after testing a large amount of items, they decide to only test those that a sick patient reported eating. Therefore, according to this model, students engage in self-regulated learning by using strategies related to monitoring and control, which allows them to actively pursue their goals and plans for accomplishing the task they are given.

This model is also particularly applicable to this study because it is the only model that views SRL as an event that unfolds in real time (Azevedo et al., 2010; Winne & Perry, 2000). For this study, we applied a sequence mining approach to examine specific events of food testing behaviors during learning via gameplay with Crystal Island. As such, we defined each testing event as an activity involving SRL and scientific reasoning, and examined sequences of how participants tested hypotheses to solve the mystery within the game.

When we study how learners use SRL strategies during learning, we should always include the context; i.e., what is required to complete the task itself (e.g., problem solving, scientific reasoning, etc.). With some GBLEs, learners must engage in scientific reasoning (or scientific inquiry), which involves using both theoretical and empirical bases for forming hypotheses that test science-related phenomena (White, Frederiksen, & Collins, 2009). As such, we used theories of SRL and scientific inquiry to investigate gameplay behaviors while learning during gameplay with Crystal Island, a GBLE that fosters SRL and scientific reasoning during learning about microbiology. For this study, we classified effective SRL as a strategic behavior, and therefore throughout the article, we refer to participants who are strategic or not strategic, which relates to their effective use of SRL processes.

1.2. Related work: research on SRL and GBLEs

Research on GBLEs has revealed not only that games are effective for learning, but has also provided guidelines for when games are the most effective. Mayer (2014) conducted a meta-analysis to investigate research comparing using games for teaching with teaching using traditional media devices (e.g., PowerPoint), as he states that research has shown that using games for teaching can be more effective. In doing so, Mayer took a four-step approach, where he did literature searches for the relevant papers, selected which papers fit the criteria for the meta-analysis, coded the experiments, and interpreted the results. Therefore, Mayer (2014) conducted this meta-analysis where he investigated different aspects (e.g., age group, content or subject, and type of GBLE) of GBLEs, and how these different types of GBLEs were found to impact learning, compared to traditional methods using media. Specifically, Mayer found that learning with games had the highest effect sizes for science and second language learning, whereas learning with games in math and language arts were found to be no better than using traditional teaching approaches (Mayer, 2014). In addition, adventure games were found to have the highest positive effects ($d = 0.72$), followed by simulations ($d = 0.62$), and quiz or puzzle games ($d = 0.45$); and games had the highest positive effects for adults or college students ($d = 0.74$), followed by secondary students ($d = 0.58$), and elementary students ($d = 0.45$). Therefore, based on this meta-analysis there is much promise for implementing games in classrooms using different domains and age groups.

Sequence mining is becoming an increasingly valuable analytical tool for assessing how students learn with ALTs, as during learning students can engage in multiple SRL strategies, and we seek to determine how their SRL unfolds over time. Studies using this approach have investigated overall performance (e.g., Baker & Corbett, 2014; Kinnebrew, Loretz, & Biswas, 2013), affect (e.g., Andres et al., 2015), and overall use of SRL skills (e.g., Bannert, Reimann, & Sonnenberg, 2014; Bouchet, Harley, Trevors, & Azevedo, 2013) during learning with various types of ALTs. Although the abovementioned studies have revealed the effectiveness of GBLEs for learning and SRL, few studies have aimed to use sequence mining to integrate how participants' scientific reasoning and inquiry, along with their metacognitive monitoring of SRL processes impacts their effectiveness in completing the games they are playing.

In addition to examining the processes of how students use SRL strategies, we must also examine *how efficiently* these processes are being used for a given task. Specifically, if students are told the overall goal of the game is to solve the mystery correctly, they might not feel it necessary to read all book content, especially content that will not be helpful for solving the mystery. In this case, the post-test might reveal a low score, however if the student solved the mystery correctly after one attempt, this can be indicative of efficient