Using Multi-Level Modeling with Eye-Tracking Data to Predict Metacognitive Monitoring and Self-Regulated Learning with CRYSTAL ISLAND

Michelle Taub[1], Nicholas V. Mudrick[1], Roger Azevedo[1], Garrett C. Millar[1], Jonathan Rowe[2], and James Lester[2]
[1] North Carolina State University, Department of Psychology, Raleigh, NC, USA
[2] North Carolina State University, Department of Computer Science, Raleigh, NC, USA

(mtaub, nvmudric, razeved, gcmillar, jprowe, lester)@ncsu.edu

Abstract. Studies investigating the effectiveness of game-based learning environments (GBLEs) have reported the effectiveness of these environments on learning and retention. However, there is limited research on using eye-tracking data to investigate metacognitive monitoring with GBLEs. We report on a study that investigated how college students' eye tracking behavior (n = 25) predicted performance on embedded assessments within the CRYSTAL ISLAND GBLE. Results revealed that the number of books, proportion of fixations on book and article content, and proportion of fixations on concept matrices—embedded assessments associated with each in-game book and article—significantly predicted the number of concept matrix attempts. These findings suggest that participants strategized when reading book and article content and completing assessments, which led to better performance. Implications for designing adaptive GBLEs include adapting to individual student needs based on eye-tracking behavior in order to foster efficient completion of in-game embedded assessments.

Keywords: Metacognition, Self-Regulated Learning, Game-Based Learning, Eye Tracking, Process Data, Scientific Reasoning.

1 SRL, Metacognitive Monitoring, and Game-Based Learning

Research on self-regulated learning (SRL) has revealed that processes related to metacognitive monitoring and control are effective for learning with advanced learning technologies, such as intelligent tutoring systems (ITSs) and game-based learning environments (GBLEs) [1]. GBLEs have been shown to be effective for learning complex topics during gameplay [2] while keeping students engaged in a learning task, particularly when designed to foster various aspects of SRL [3]. These environments have been developed to afford opportunities to engage in scientific reasoning and problem solving [4], and studies have found that GBLEs are often more effective than traditional teaching methods, in terms of learning and retention [6].

adfa, p. 1, 2011.

Despite the growing evidence indicating that GBLEs lead to improved learning outcomes [5], prior research on metacognitive monitoring and SRL within GBLEs has primarily focused on classifying SRL behaviors and relating them to in-game behavior and learning outcomes [3, 6]. In this paper, we aim to integrate how we can use trace data to track students' metacognitive monitoring and SRL to assess performance on concept matrices, an in-game embedded assessment tool within CRYSTAL ISLAND. We investigated if students were using metacognitive monitoring strategies, as indicated by their log file and eye tracking behavior during knowledge construction activities (e.g., reading) related to scientific reasoning to perform successfully on embedded measures of text comprehension (e.g., completing in-game concept matrices), as evidence of SRL and scientific reasoning in GBLEs. Multi-level modeling (MLM) is an ideal analytical technique to assess student learning with GBLEs because it enables statistical analyses of learning events at nested levels of abstraction that do not require restrictive statistical assumptions [7].

2 Method

35[1] undergraduate students from North Carolina State University (50% female), with ages ranging from 18 to 29 (M = 20.18, SD = 2.38), participated in the study. Prior to beginning the study, the students were randomly assigned to one of three experimental conditions. Students were compensated for their participation in the study, receiving \$10 per hour, up to a total of \$30 for full participation.

CRYSTAL ISLAND is a 3D game-based learning environment designed to foster students' self-regulated learning, problem solving, scientific reasoning, and literacy skills [8]. When participants begin to play CRYSTAL ISLAND, they are informed of an outbreak that has impacted a group of scientists on the remote island. The student's task is to identify the epidemic that has spread amongst the scientists, determine the disease's transmission source, and recommend a treatment and prevention plan for the island's inhabitants. To do so, participants explore the virtual environment from a first-person perspective, navigating between five different buildings on the island: an infirmary, a living quarters, a dining hall, a laboratory, and the lead scientist's residence. These activities contribute to students' engaging in scientific reasoning, which involves hypothesis generation and testing, followed by forming conclusions based upon gathered test results [4]. Participants engaged in scientific reasoning when playing CRYSTAL ISLAND; they generated hypotheses based on the clues they gathered from non-player characters, reading books and articles, viewing posters, and testing their hypotheses about the spreading disease's transmission source. In order to complete the game, participants must submit a correct diagnosis.

One assessment tool, the concept matrix, was embedded into gameplay, such that there was a concept matrix to complete with every book or research paper the participant read (Fig. 1.). The matrices contained questions regarding the book content in

[1] Data from 25 participants were included in this analysis because the other participants were in the *No Agency* condition (see Results section below).

multiple-choice format. Participants were not restricted to answer the questions without returning to the text (i.e., they did not have to memorize the content). In addition, participants were given three attempts at completing the questions in the concept matrix, and if they failed to answer the questions correctly after three attempts, the game auto-filled the responses for participants (to ensure that they were eventually provided with the correct answer and were given potential information needed to help solve the mystery). Concept matrices are used to assess students' understanding of scientific concepts introduced in the reading material within the game.

When students played CRYSTAL ISLAND, we collected multi-channel SRL process data, including (1) software log files and (2) eye-tracking data. The log-file data captured student i nteractions with the game environment, including timestamp, action type, location, object, and characters involved in the interaction. The eye-tracking data provided gaze patterns and fixation behaviors on predefined areas of interest (AOIs) in the game, such as fixation duration on book content and fixation duration on concept matrices. To code and score the data, the number of concept matrix submission attempts (dependent variable) was calculated from the software log data capturing these events (M = 1.03, SD = .75, across all three conditions). We used three predictor variables for this analysis. The number of books and articles read was extracted from the log files. This variable was calculated based upon the total number of books that participants selected throughout gameplay (M = 24.57, SD = 8.57, across all three conditions). The other two variables were extracted from the eye-tracking data: (1) the proportion of time fixating on book content, and (2) the proportion of time fixating on an associated concept matrix. These variables were calculated by dividing the fixation duration of each activity over the total book fixation duration, yielding one proportion for fixation duration on book content (M = .33, SD = .22, across all three conditions), and one proportion for fixation duration on book concept matrices (M = .19, SD = .13, across all three conditions). Once calculated, these data were used to address the research questions posed for this analysis.

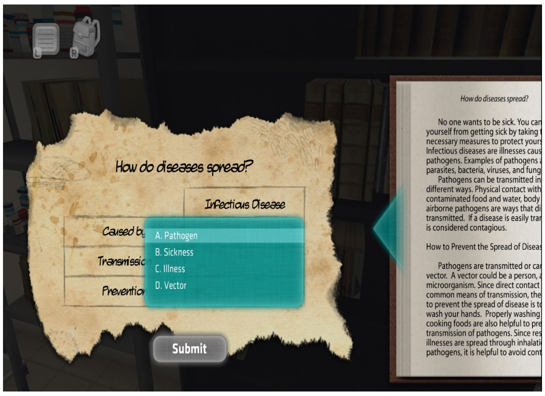

Fig. 1. Screenshot of a virtual book and associated concept matrix in CRYSTAL ISLAND.

3 Results

Prior to gameplay, participants were randomly assigned to 1 of 3 experimental conditions (*No Agency, Partial Agency,* or *Full* Agency). However, for this study, we included data from only the 2 interactive conditions, because the *No Agency* condition did not allow participants to select books to read, nor did participants in this condition complete concept matrices; students simply watched an expert player perform these activities. Thus, we only analyzed data from 25 participants, with n = 12 for the *Full Agency*, and n = 13 for the *Partial Agency* condition.

For this study we used multi-level modeling [7]. We ran three separate models, each with the same dependent variable: the number of concept matrix attempts. This required only one fully unconditional model to be run. Results from the fully unconditional model revealed there was significant between-subjects (τ_{00} = .03, z = 2.01, p = .02) and within-subjects (σ^2 = .52, z = 16.10, p < .0001) variance in the number of concept matrix attempts, with 5.6% variance between-subjects, and 94.4% variance within-subjects. Thus, this model indicated that it was appropriate to continue to run models with predictor variables, as was done for the following research questions.

3.1 Research Question 1: Is there an association between the number of books read and the number of concept matrix attempts?

To address this research question, we ran a means-as-outcomes regression model with constrained slopes, with the number of books as the predictor variable (between-subjects, level 2) and the number of concept matrix attempts as the dependent variable. Results indicated that an increase in the number of books was associated with a decrease in the number of concept matrix attempts; γ_{10} = -.02, t = -5.56, p < .0001. This model explained 100% of the between-subjects variance in number of concept matrix attempts. In general, this finding indicates that as participants were selecting more books to read, they were making fewer concept matrix attempts, indicating that as they were reading more books, they were performing better on the concept matrices associated with each book.

3.2 Research Question 2: Is there a relationship between concept matrix attempts and proportion of fixations on book content, and does this relationship depend on the proportion of fixations on book concept matrices?

For this research question, we ran a level 1 moderation model with constrained slopes, with concept matrix attempts as the dependent variable, and the proportion of fixations on book content and the proportion of fixations on book concept matrices as the predictor variables (within-subjects, level 1). Results indicated that the proportion of fixations on book content was not associated with concept matrix attempts (γ_{10} = .06, t = .27, p = .79), nor was there an association between the proportion of fixations on book concept matrices and concept matrix attempts (γ_{20} = .07, t = .21, p = .84). However, there was a significant interaction; γ_{30} = 8.03, t = 6.92, p < .0001, such that participants with the fewest concept matrix attempts had the lowest proportion of fixations on book content, as well as on book concept matrices. This model explained 18.1% of the within-person variance in concept matrix attempts. This finding indicates that a lower amount of fixation durations on both book content and concept matrices resulted in better performance on the concept matrices, such that spending more time reading the content and concept matrices did not result in better performance on the matrices.

3.3 Research Question 3: Does the relationship between concept matrix attempts and number of books read depend on the proportion of fixations on book content and on the proportion of fixations on book concept matrices?

This final research question used a 3-way cross-level interaction model with constrained slopes, with concept matrix attempts as the dependent variable and all three predictor variables used in the previous analyses (i.e., number of books – level 2 variable, proportions of fixations on book content and book concept matrices – level 1 variables). Results revealed a significant 3-way cross-level interaction; $\gamma_{31} = .26$, $t = 2.16$, $p = .03$, such that participants who had the least amount of concept matrix attempts read more books and had lower proportions of fixations on book content (fig. 2, left) and on book concept matrices (fig. 2, right). This model accounted for 19.3% of the within-person variance in concept matrix attempts. Overall, these findings reveal that reading more books led to better performance on the concept matrices, however this was in combination with spending less time reading the content and concept matrices associated with each book.

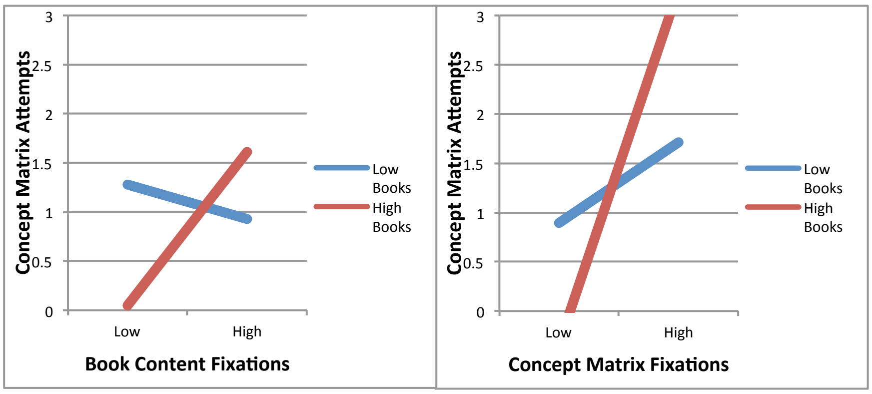

Fig. 2. Interaction between fixations on book content and number of books (left) and book concept matrices and number of books (right), each on the number of concept matrix attempts.

4 Conclusions and Future Work

In this study, we used MLM to explore the links between theory (SRL and metacognitive monitoring), data channels (eye tracking and log files), and performance on an in-game assessment tool to examine how students used cognitive and metacognitive processes (reading comprehension) during knowledge construction activities related to scientific reasoning (completing the concept matrix), to provide evidence of SRL and scientific reasoning with the CRYSTAL ISLAND GBLE. Results indicated that these activities did significantly predict the number of concept matrix attempts, such that selecting more books was associated with fewer matrix attempts. However, fixating on more book content and fixating on more concept matrix content was associated with more matrix attempts, with fewer attempts as a more desirable outcome.

From this analysis, we cannot determine the sequence of events, such that we cannot confirm participants were transitioning from looking at the specific questions in the matrix, and finding those responses in particular areas within the text. Therefore, we cannot conclude that students were engaging in strategic reading, based on accurate monitoring, however, future studies will investigate the sequential order of reading books and completing their concept matrices in order to test this hypothesis. In particular, the use of analytical techniques that are amenable to sequence data show especial promise, such as sequence mining [9].

These results have important implications for designing intelligent GBLEs that afford students the opportunities to engage in cognitive, affective, metacognitive, and motivational processes to foster learning and scientific reasoning. Additionally, including adaptive scaffolding can improve the success of these environments in fostering learning during gameplay, such that they can use eye tracking to provide tailored scaffolding based on student strategy use. Improving the intelligence and efficiency of GBLEs can be beneficial to ensure that each student's real-time cognitive and metacognitive learning needs are being met, while still enjoying learning during gameplay.

Acknowledgments

This study was supported by funding from the Social Sciences and Humanities Research Council of Canada. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Social Sciences and Humanities Research Council of Canada.

References

1. Azevedo, R., & Aleven, V. (Eds.). (2013). International handbook of metacognition and learning technologies. Amsterdam, The Netherlands: Springer.
2. Lester, J., et al. (2013). Serious games get smart: Intelligent game-based learning environments. AI Magazine, 34, 31-45.
3. Sabourin, J. L., & Lester, J. C. (2014). Affect and engagement in game-based learning environments. IEEE Transactions on Affective Computing, 5, 45-56.
4. Spires, H. A., et al. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. Journal of Educational Computing Research, 44, 453-472.
5. Wouters, P., et al. (2013). A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology, 105, 249-265.
6. Rodeghero, P., et al.. (2014). Improving automated source code summarization via an eye-tracking study of programmers. In Proceedings of the 36th International Conference on Software Engineering 2014 (pp. 390-401). New York, NY: ACM.
7. Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.
8. Rowe, J. P., et al. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. International Journal of Artificial Intelligence in Education, 21, 115-133.
9. Kinnebrew, J. S., et al. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. Metacognition and Learning, 9, 187-215.