EXAMINING THE TIMING OF METACOGNITIVE MONITORING JUDGMENTS IN A GAME-BASED LEARNING ENVIRONMENT

John L. Nietfeld[1], Samira Syal[1], Rayne A. Sperling[2]

[1]North Carolina State University (USA)
[2]The Pennsylvania State University (USA)

Abstract

The purpose of the current study was to analyze the impact of delayed monitoring judgments on both monitoring accuracy and science knowledge in a game-based learning environment called MISSING MONTY. Fifth-grade students from public schools in the USA were randomly assigned to either an immediate monitoring (IM) (*n* = 142) condition or to a delayed monitoring (DM) condition (*n* = 171). All students completed a pre and posttest of science knowledge and made item-level confidence judgments on each test. The students then played MISSING MONTY for approximately 2-5 weeks depending upon class schedule. During gameplay students visited various animal researchers, read informational texts, and completed knowledge and monitoring challenges. In the IM condition, students rated their confidence on a 100-point scale immediately following each item. In the DM condition, the students first completed the knowledge challenge and then provided monitoring judgments following the completion of all items. Results showed significant improvements for science knowledge and monitoring accuracy for both groups, however no significant differences were found between the two conditions Thus, MISSING MONTY appeared to have positive effects on both resultant science knowledge and monitoring accuracy regardless of when monitoring was assessed. Implications for the design of learning environments and SRL will be discussed.

Keywords: Metacognition, monitoring, game-based learning environment.

1 INTRODUCTION

Metacognitive monitoring is a critical component of self-regulated learning (SRL) [1] and accurate monitoring is critical for performance in game-based learning environments (GBLEs) [2]. Examples of effective monitoring in educational contexts involve having an accurate understanding of one's knowledge, being able to regulate cognitive strategies, and being able to dynamically assess what information needs to be restudied, just to name a few. A rather large literature has emerged in the past 40 years examining how metacognitive monitoring occurs, how to measure monitoring, the relationship between metacognitive processes and learning performance outcomes, and more recently in developing techniques to improve monitoring accuracy.

There is a consistent positive relationship between accurate monitoring and retention and achievement [3,4]. Delving more deeply, studies are now attempting to understand how best to improve monitoring accuracy. Research, primarily from lab studies using paired associate tasks, have generally found that having learners make delayed judgments of learning (JOLs) rather than immediate JOLs improves monitoring accuracy [5,6]. Yet, while monitoring accuracy improves with delayed JOLs the effect of the delay on memory-based performance is quite small [6]. In fact, evidence now suggests that the assumption of learners engaging in 'covert retrieval' when asked to produce delayed judgments fares less well than having learners make explicit attempts at retrieval [7], likely because learners truncate their retrieval attempts when making their judgments.

However, less is known about the impact of delayed monitoring judgments on monitoring accuracy and memory performance within externally-valid learning environments. Moreover, most studies of delayed JOLs have relied upon relative indices (e.g., Gamma) to measure accuracy rather than absolute accuracy that compares a judgment to a criterion task [8]. The purpose of the current study was to investigate change over time as the results of activities using a digital GBLE, therefore we chose to measure monitoring accuracy with absolute measures.

A great deal of research has been invested in attempting to curb learners' overconfidence [9]. Prior attempts to improve monitoring accuracy in classroom environments has been challenging, many times revealing no improvements or sustained overconfidence even after repeated attempts of providing confidence judgments over time [10,11]. However, findings are emerging that suggest pathways that

lead to improvements. Important factors include modeling effecting monitoring processes [12], providing numerous practice opportunities distributed over time combined with training and individual feedback [13,4], and providing monitoring exercises with customized feedback in online environments [14]. An attempt was made to provide training, customized feedback, and distributed practice over time in the GBLE used in the current study.

MISSIONS WITH MONTY is a GBLE focused on improving 5th graders' science literacy. More specifically the program targets metacomprehension skills for informational texts aligned with classroom science curriculum. The program, funded by the National Science Foundation, promotes SRL and includes ecosystem curriculum. In the current study, we examined the timing of metacognitive monitoring judgments in the first MISSIONS WITH MONTY module entitled MISSING MONTY. For this module the player fills the role of a promising young science professor traveling to work with Monty, a monitor lizard and world-renowned scientist known for his ability to solve real-life problems. Monty has created Wildlife University (WU) in a remote rainforest. The students and professors at WU are animals of many different types focused on becoming more scientifically literate in order to save their natural habitats. Unfortunately, upon arrival at WU the player is presented with two major problems to solve: 1) Monty has gone missing and 2) WU has been recently closed due to animals getting sick. Students then proceed to take on the role of a researcher seeking to solve these two overarching problems by collecting key information from various animal researcher sites. Eventually, students narrow their information on a detective board and present their final hypothesis for the sickness.

The primary purpose of the current study was to examine the extent to which the timing of monitoring judgments impacted the accuracy of such judgments. Secondarily, we tested the extent to which MISSING MONTY was successful in promoting science content knowledge.

2 METHODOLOGY

Fifth-grade students (Mean Age = 10.68 years; 50.6 % girls; 46.2% boys) from public schools in North Carolina, USA participated as part of regular classroom instruction and were randomly assigned to either the immediate monitoring (IM) (n = 142) or the delayed monitoring (DM) condition (n = 171).

All students completed a pre and posttest of science knowledge and provided item-level confidence judgments on each test. The test consisted of 20 multiple-choice items that were reviewed by teachers and further developed in a validation study. The content was aligned with the passages in the GBLE along with the state curriculum and included both declarative and conceptual level items.

Students first completed the 20-item science knowledge test in addition to some other demographic and motivation questionnaires. The following week the students then began playing MISSING MONTY for approximately 2-5 weeks depending upon class schedule. Students navigated through the GBLE individually with the teacher's role only to intervene for technical issues. In the week following gameplay students completed the science knowledge test again as a post measure. MISSING MONTY allows students to read texts and complete challenges at their own pace and as a result, students' completion status varied.

During gameplay students visited various animal researchers, read informational texts (see Figure 1), and completed knowledge and monitoring challenges. They attempted to determine the source of the illness at WU by saving critical information and understanding gained from the text passages presented by the animal researchers. Periodically, they were also challenged to display their multiple-source understanding within the game context. Students received badges depending upon their performance levels for the in-game challenges.

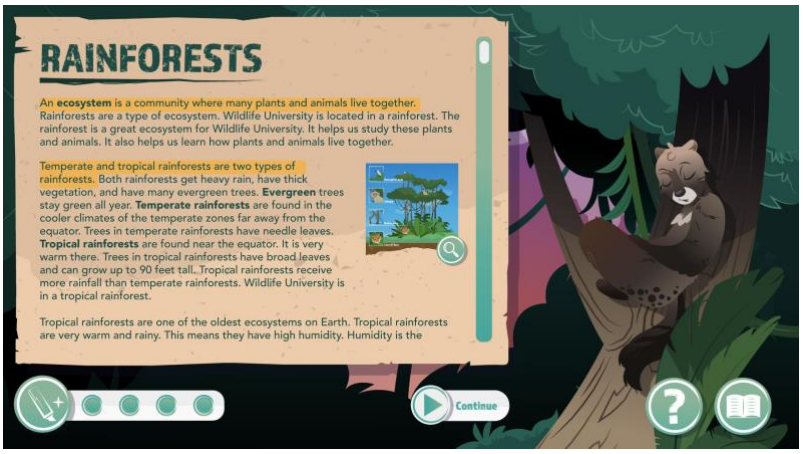

Figure 1. Text interaction in MISSING MONTY.

As students progressed through the GBLE they responded to multiple-choice items within the context of "Knowledge Challenges" following each text that was encountered. In the Immediate Monitoring (IM) condition, students rated their confidence on a 100-point scale immediately following each item presented in the Knowledge Challenge. The confidence estimate was displayed in the form of a rainforest vine to fit within the game narrative (see Figure 2) with immediate feedback provided (see Figure 3). In the Delayed Monitoring (DM) condition, the students first completed all five Knowledge Challenge items and then provided each their monitoring judgments in succession in a "Monitoring Challenge" following the Knowledge Challenge. MISSING MONTY included 16 passages spread over four sections or "Days." Of the total 16 passages 3 had 9 items each because they contained graphs. These passages were in Days 3 and 4.

Metacognitive monitoring was derived as a measure of calibration by comparing the student's judgment for each item to their performance. Performance was scored as either a 1 "correct" or 0 "incorrect." The calibration variable consisted of the absolute value of the difference between the confidence judgment and performance for each test item, summed over all items on a test and divided by the total number of items. Scores could range from zero (perfect calibration) to one (complete lack of calibration). For example, if a confidence rating for a given item was 79 and the participant answered the question correctly the accuracy score for that item would be 0.21 (absolute value of 1– 0.79).

Figure 2. Confidence estimate in MISSING MONTY.

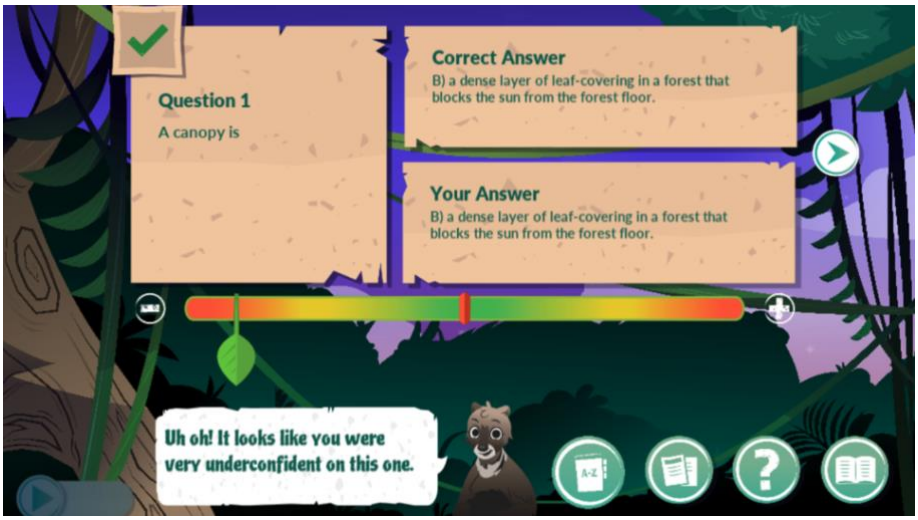

Figure 3. Confidence judgment feedback in MISSING MONTY.

3 RESULTS

Descriptive statistics for the primary study variables are presented in Table 1. Results showed no significant difference between the conditions on science knowledge [$F(1, 113) = 0.61$, $p = .519$]. However, both groups showed significant improvements in science knowledge [$F(1, 113) = 55.18$, $p < .001$, $\eta_p^2 = .33$] from pre to posttest (see Figure 4). In addition, results indicated that there was no significant difference between conditions on monitoring accuracy (calibration) [$F(1, 113) = 1.276$, $p = .261$]. However, both groups showed significant improvements in calibration accuracy [$F(1, 113) = 23.45$, $p < .001$, $\eta_p^2 = .17$) (see Figure 5). Thus, MISSING MONTY appeared to have positive effects on both resultant science knowledge and monitoring accuracy regardless of when monitoring was assessed.

Table 1. Means and standard deviations for the study variables.

	Science Content Knowledge				Monitoring Accuracy			
	Pretest		Posttest		Pretest		Posttest	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Immediate Monitoring	12.81	3.42	14.63	3.68	.34	.11	.29	.13
Delayed Monitoring	12.43	3.87	14.14	4.27	.34	.11	.30	.15

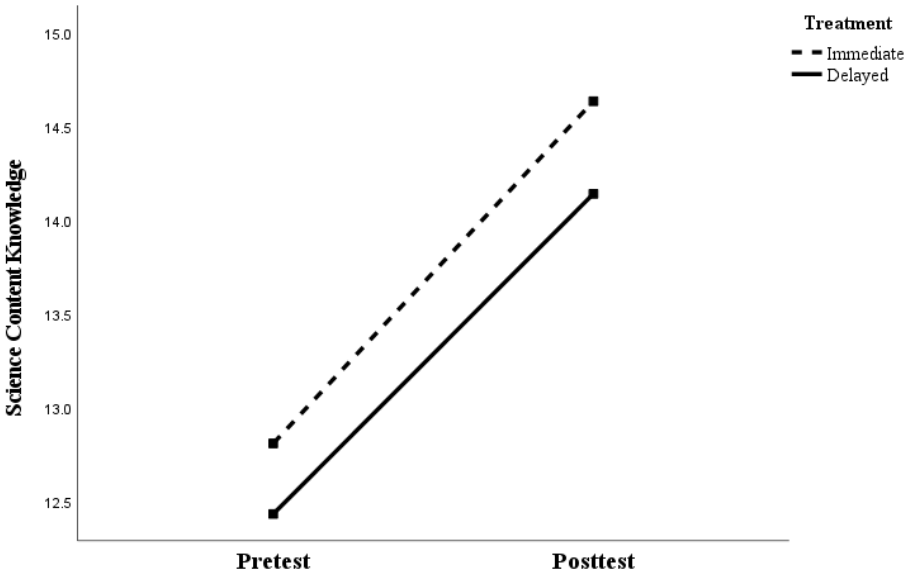

Figure 4. Pre to Posttest Science Knowledge Scores by Group.

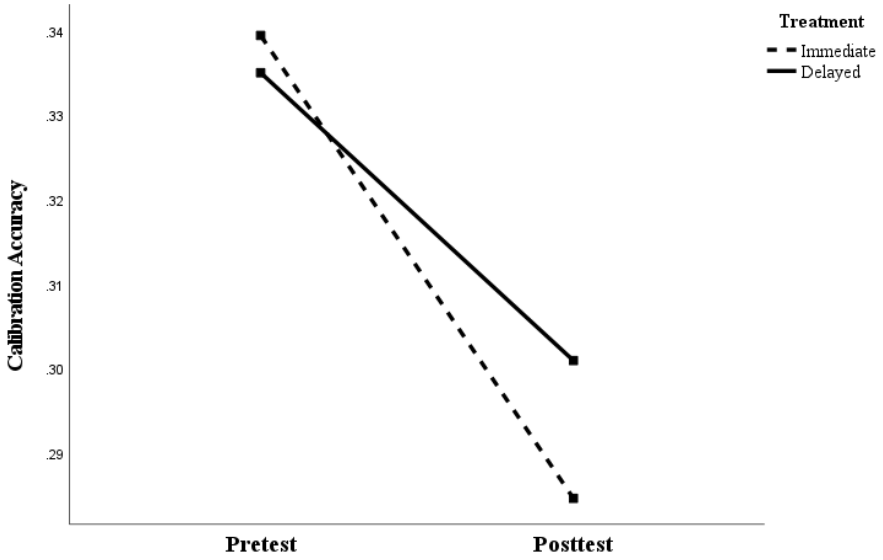

Figure 5. Pre to Posttest Monitoring Accuracy by Group.

4 CONCLUSIONS

This study examined the impact of the timing of metacognitive confidence judgments upon monitoring accuracy and science content knowledge. Research in externally valid educational settings is lacking in this area. Prior work has reported advantages with regard to monitoring accuracy for delayed judgments [6], however negligible impacts for memory performance. In addition, changes in monitoring accuracy in classroom-based contexts is typically difficult to obtain, requiring distributed practice over time along with specific feedback.

In the current study the MISSING MONTY GBLE was integrated within typical classroom instruction and students were provided numerous opportunities to practice making judgments distributed across time while simultaneously being provided with customized feedback. Students subsequently improved their monitoring accuracy and science knowledge across both conditions. Thus, the delay in providing the confidence estimate did not appear to influence judgment accuracy or science content knowledge. However, it appears that the MISSING MONTY program has efficacy to promote monitoring accuracy regardless of the timing of the judgment, likely due to the ability of the program to give clear and timely

feedback customized for each learner. Moreover, the monitoring practice and training is conducted in a game-based environment that may increase the engagement or willingness of the learner to provide conscientious judgments, yet this should be further investigated given the unclear impact of GBLEs on motivation [15].

These results are very encouraging from the perspective of promoting SRL in online environments. However, further investigations are needed including those that consider comparison conditions that lack monitoring judgments and/or feedback to isolate effects. In addition, the current framework should be tested on changes that take place on judgments within the GBLE itself and on other forms of assessment outside the GBLE as well as within different GBLEs altogether.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation [award #DRL-1761110].

REFERENCES

[1] P. H., Winne, A. F. Hadwin, "Studying as Self-Regulated Learning" in *Metacognition in Educational Theory and Practice* (D. J. Hacker, J. Dunlosky, A. C. Graesser eds.), pp. 293-310., New York: Springer, 1998.

[2] J. L. Nietfeld, J., L. R. Shores, K. F. Hoffmann, "Self-regulation and gender within a game-based learning environment," *Journal of Educational Psychology*, vol. 106, pp. 961-973, 2014.

[3] J. Dunlosky, K. A. Rawson, "Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention," *Learning and Instruction*, vol. 22, pp.271-280, 2012.

[4] J. L. Nietfeld, L. Cao, J. Osborne, "The effect of distributed monitoring exercises and feedback on performance and monitoring accuracy," *Metacognition and Learning*, vol. 2, pp. 159-179, 2006.

[5] J. Dunlosky, T. O. Nelson, "Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect," *Memory & Cognition*, vol. 20, pp. 373-380, 1992.

[6] M. G. Rhodes, S. K. Tauber, "The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review," *Psychological Bulletin*, vol. 137, no. 1, pp. 131-148, 2011.

[7] E. Tekin, H. L. Roediger III, "The Effect of Delayed Judgments of Learning on Retention," *Metacognition and Learning*, vol. 16, pp. 407-429, 2021.

[8] G. Schraw, "Measures of feeling-of-knowing accuracy: A new look at an old problem," *Applied Cognitive Psychology*, vol. 9, pp. 321-332, 1995.

[9] D. J. Hacker, L. Bol, D. D. Horgan, E. A. Rakow, "Test prediction and performance in a classroom context," *Journal of Educational Psychology*, vol. 92, no. 1, pp. 160-170, 2000.

[10] J. L. Nietfeld, L. Cao, J. Osborne, "Metacognitive monitoring accuracy and student performance in the classroom," *Journal of Experimental Education*, vol. 74, pp. 7-28, 2005.

[11] N. L. Foster, C. A. Was, J. Dunlosky, R. M. Isaacson, "Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions," *Metacognition and Learning*, vol. 12, pp. 1-19, 2017.

[12] J. D. Huff, J. L. Nietfeld, "Using strategy instruction and confidence estimates to improve metacognitive monitoring skills," *Metacognition and Learning*, vol. 4, pp. 161-176.

[13] M. Händel, B. Harder, M. Dresel, "Enhanced monitoring accuracy and test performance: Incremental effects of judgment training over and above repeated testing," *Learning and Instruction*, vol. 65, 2020, Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0959475218308788?via%3Dihub

[14] A. B. H. de Bruin, E. M. Kok, J. Lobbestael, A. de Grip, "The impact of an online tool for monitoring and regulating learning at university: Overconfidence, learning strategy, and personality," *Metacognition and Learning*, vol. 12, pp. 21-43, 2017.

[15] P. Wouters, C. van Nimwegen, H. Oostendorp, E. D. van der Spek, “A meta-analysis of the cognitive and motivational effects of serious games,” *Journal of Educational Psychology*, vol. 105, pp. 249-265, 2013.