Metacognitive Prompt Overdose: Positive and Negative Effects of Prompts in iSTART

Kathryn S. McCarthy, Amy M. Johnson, Aaron D. Likens, Zachary Martin, Danielle S. McNamara
Arizona State University
Tempe, AZ, USA
{ksmccar1, amjohn43, alikens, zsmartin, dsmcnama}@asu.edu

McCarthy, K. S., Johnson, A. M., Likens, A. D., Martin, Z., & McNamara, D. S. (2017). Metacognitive prompt overdose: Positive and negative effects of prompts in iSTART. In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.), *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 404-405). Wuhan, China: International Educational Data Mining Society. Published with acknowledgment of federal support.

Author's Note

This research was supported in part by IES Grant R305A130124. Opinions, conclusions, or recommendations do not necessarily reflect the views of the IES.

Metacognitive Prompt Overdose: Positive and Negative Effects of Prompts in iSTART

Kathryn S. McCarthy, Amy M. Johnson, Aaron D. Likens, Zachary Martin, Danielle S. McNamara
Arizona State University
Tempe, AZ, USA
{ksmccar1, ambw, alikens, zsmartin, dsmcnama}@asu.edu

ABSTRACT
Interactive Strategy Training for Active Reading and Thinking (iSTART) is an intelligent tutoring system that supports reading comprehension through self-explanation (SE) training. This study tested how two metacognitive features, presented in a 2 x 2 design, affected students' SE scores during training. The *performance notification* feature notified students when their average SE score dropped below an experimenter-set threshold. The *self-rating* feature asked participants to rate their own SE scores. Analyses of SE scores during training indicated that neither feature increased SE scores and, on the contrary, seemed to decrease SE performance after the first instance. These findings suggest that too many metacognitive prompts can be detrimental, particularly in a system that provides metacognitive strategy training.

Keywords
intelligent tutoring systems; metacognition; educational games; system interaction logs

1. INTRODUCTION
Intelligent tutoring systems (ITSs) provide an opportunity for extended training and individualized feedback to support the development of skills and strategies. One such ITS, Interactive Strategy Training for Active Reading and Thinking (iSTART) uses self-explanation (SE) training as a means of increasing students' comprehension of complex texts [4]. iSTART provides instruction on SE strategies through lesson videos, guided demonstration, and practice. Research indicates that prompting metacognition, or reflection on one's own knowledge, can enhance the benefits of training within computer-based learning [1]. In this study, we expand upon previous research to investigate how two metacognitive features affect the SE scores during iSTART practice.

In iSTART's generative practice, students write their own SEs and a natural language processing (NLP) algorithm immediately provides a score of poor (0), fair (1), good (2), or great (3). The two metacognitive features were implemented within this generative practice. The first feature is a *performance notification* that alerts students that their SE score is below 2.0 and sends them to Coached Practice for remediation. The second feature is a *self-rating* that prompts students to rate the quality of their SE before receiving the computer-generated score. The performance notification encourages metacognition indirectly, whereas the self-rating is a direct metacognitive prompt [6]. The current study expands on data reported in [3], which further demonstrated the positive effects of iSTART on deep comprehension, but also indicated that neither metacognitive feature affected post-training learning outcomes. In this study, we explore the log-data to investigate how these two metacognitive features, both individually and in combination, affect SE scores during iSTART generative practice.

Based on previous work [6], we predicted that the performance notification would increase SE scores immediately after the first instance of the notification. In [6], however, the instruction was brief, and did not allow examining further instances of the notification. In this study, we examine the effects of the notification after the initial instance during a longer duration study. Consistent with previous research [5], we had predicted that self-ratings would improve performance. Of particular interest was the interaction of the two features. One hypothesis is that there would be an additive effect such that having both features would yield the greatest SE score improvement [2]. An alternative hypothesis is that the redundancy of the two features would result in an interactive, and possibly negative effect [4].

2. METHODS
2.1 Participants
As part of the larger study reported in [3], 116 high school students (M_{age}=17.67, SD=1.30) received monetary compensation for their participation.

2.2 Design and procedure
The study employed a 2(performance notification: off, on) x 2(self-rating: off, on) between-subjects design. Participants completed iSTART training in three 2-hour sessions. Participants first watched iSTART video lessons that provide instruction on the purpose of SE training and five comprehension strategies (comprehension monitoring, paraphrasing, prediction, elaboration, and bridging). Next, participants completed one round of Coached Practice, in which a pedagogical agent provides individualized feedback on students' self-explanations. Participants were then allowed to move freely throughout the system to interact with videos, Coached Practice, identification games, and generative games for the remainder of the training sessions. The metacognitive features were implemented only during generative games. Performance notifications were triggered each time the average SE score was less than 2.0 and self-rating prompts were

triggered on randomly-determined self-explanations approximately 1/3 of the time.

3. RESULTS

We calculated a *gain score* to compare the average SE score in the game before and immediately following an average generative game score of 2.0 indicative of when the performance notification was triggered (or *would have triggered* in the notification off conditions). We used log-data to identify participants who completed at least one game in which their average SE score was less than 2.0 (n=78). Though the performance notification could be triggered as many times as necessary, most participants had no more than two instances of less than 2.0 average SE scores (Fig. 1). As participants were able to move freely through the system, only 48 participants (across all conditions) followed the *generative game, notification, generative game* needed to calculate a gain score. These participants were relatively evenly distributed across the conditions. We analyzed the first two instances of average SE scores less than 2.0 for these 48 participants.

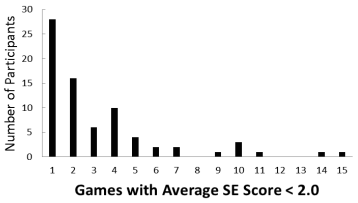

Figure. 1 *Frequency of Games with Average SE Scores < 2.0*

For the first instance of notification, the average gain scores in all conditions were positive. Though the pattern of gain scores for the performance notification is consistent with previous findings [3], an ANOVA indicated no effect of notification, no self-rating, and no interaction, all $F(1, 47) < 2.00$ (Fig. 2, *left*).

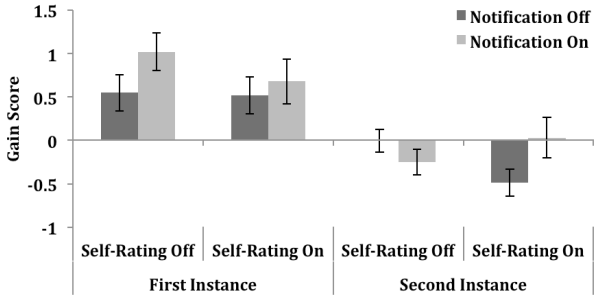

Figure 2. *Gain score in 1st and 2nd instance of avg. SE score < 2.0 as a function of performance notification and self-rating*

Fewer participants (n=27) had a second instance of notification. Contrary to the scores following the first instance, in this second instance, average gain scores were either near zero or negative, indicating that the scores after notification were the same or lower than before the notification. An ANOVA revealed no main effect of performance notification or self-rating, Fs < 1.00, *ns.* There was a significant notification by self-rating interaction indicating that having neither feature or both features did not affect SE score, but that the presence of only one metacognitive feature was detrimental to SE score, $F(1, 26)=5.46$, $p < .05$, $\eta^2_p=.17$ (Fig. 2, *right*).

4. CONCLUSIONS

These findings indicate that neither metacognitive feature had a consistent effect on SE quality during iSTART training. Though there was an overall increase in SE score in the first instance (as indicated by positive gain scores), there was no significant effect of either performance notification or self-rating compared to control. In the second instance, the interaction should be interpreted with caution given the small sample size. Nonetheless, the features did not improve SE score, and were potentially detrimental to performance. One explanation for these findings is that iSTART intrinsically instructs on metacognitive strategies. Hence, the inclusion of additional metacognitive prompts may be redundant, if not overwhelming, at least after the first instance.

These results were not consistent with extant research, and may be particular to iSTART. Certainly further analyses and studies are merited and will be explored. Nonetheless, given that the neither prompt showed post-training learning outcomes [3] or sustained training benefits, we do not intend to include these features in future implementations of iSTART, and we would caution other researchers to consider the possibility of potential metacognitive prompt over-dosages.

5. ACKNOWLEDGMENTS

This research was supported in part by IES Grant R305A130124. Opinions, conclusions, or recommendations do not necessarily reflect the views of the IES.

6. REFERENCES

[1] Azevedo, R., Hadwin, A.F.: Scaffolding self-regulated learning and metacognition–Implications for the design of computer-based scaffolds. *Instructional Science 33*: 367-379 (2005)

[2] Flavell, J. H.: Metacognition and cognitive monitoring: A new area of cognitive– developmental inquiry. American Psychologist. 34(10), 906 (1979)

[3] McCarthy, K.S., Jacovina, M. E., Snow, E.L. Guerrero, T. A., & McNamara, D.S.. *iSTART therefore I understand: But metacognitive supports did not enhance comprehension gains.* Proceedings of the 18th *International Conference on AIED.* Wuhan, China (2017).

[4] McNamara, D.S., Levinstein, I.B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers, 36*, 222-233.

[5] Schraw, G. & Dennison, R.S.: Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*: 460-475 (1994)

[6] Snow, E.L., et al.: Promoting metacognitive awareness within a game-based intelligent tutoring system. In Mitrovic, A., Verdejo, F., Conati, C., Heffernan, N. (eds), *Proceedings of the 17th International Conference on AIED 2015.* Madrid, Spain: Springer, pp 786-789 (2015)