

Impact of Different Pedagogical Agents' Adaptive Self-regulated Prompting Strategies on Learning with MetaTutor

François Bouchet, Jason M. Harley, and Roger Azevedo

McGill University, SMART Laboratory, Montreal, Canada
francois.bouchet@mcgill.ca

Abstract. Extended interactions with a pedagogical agent (PA) assisting students to enact cognitive and metacognitive self-regulated processes requires the system to adapt the types and frequency of scaffolding. We compared learners' perception of PAs' prompts with MetaTutor, a hypermedia adaptive learning environment, with 40 undergraduates randomly assigned to one of three conditions: non-adaptive prompting (NP), frequency-based adaptive prompting (FP) and frequency and quality-based adaptive prompting (FQP). Results indicate learners are unable to reliably perceive differences in the number of prompts received, though these differences are reflected in positive outcomes in terms of SRL processes enacted and learning gains, and negative outcomes in terms of self-reported satisfaction. Preliminary results indicated that more frequent, but adaptive prompting is an efficient scaffolding strategy, despite negatively impacting learners' satisfaction.

Keywords: pedagogical agents, intelligent tutoring systems, adaptivity, user perception, self-regulated learning, metacognition.

1 Need for Adaptive Prompt Frequency

ITSs' core ability is to provide individualized instruction, feedback, and scaffolding based on a dynamic assessment of learners' emerging understanding of the content, use of learning strategies, and metacognitive judgments to help learners develop cognitive skills [1, 2]. This paper assessed the impact of different pedagogical agents' (PAs) adaptive prompting of self-regulated learning (SRL) processes during learning with MetaTutor [3], a multi-agent ITS designed to track, detect, model, and foster cognitive and metacognitive processes, in which 4 PAs help students learn about the circulatory system, using 38 pages with text and diagrams, accessible through a table of contents [3]. PAs scaffold by prompting students to engage into SRL processes, which they can also self-initiate through a palette of actions (in which case PAs simply accompany their deployment through a dialogical interaction). More specifically, we investigate: (1) how the frequency changes affect learners' use of SRL processes and (2) whether learners perceived changes in the frequency of prompts they received.

2 Method

40 undergraduates (62.5% female) were randomly assigned to 3 experimental conditions: non-adaptive prompt (NP, from a larger sample of 58), frequency-based prompt (FP) and frequency and quality-based prompt (FQP, cf. Table 1). As learners in FP and FQP were similar, they were sometimes grouped to have two samples of an identical size. In the *NP condition*, learners received a moderate, but constant amount of prompts from the PAs (~ 1 every 10 min.) to engage in SRL processes. In the *FP and FQP conditions*, they received more prompts at first (~ 3.5 every 10 min.), but the probability of each prompt category (monitoring and strategy) being triggered decreased after each received prompt and after each self-initiated enactment of an SRL process. In the *FQP condition*, the probability of each prompt category to be triggered could also increase (1) if learners did not comply to a PA's non-mandatory prompt (e.g., a suggestion to open an image), or (2) if learners' metacognitive judgment was inaccurate (e.g., evaluating a page irrelevant to the active sub-goal as relevant).

Table 1. Pre-test and post-test means and standard deviations across conditions

	NP (<i>n</i> = 20)		Original NP (<i>n</i> = 58)		FP (<i>n</i> = 8)		FQP (<i>n</i> = 12)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pre-test score (out of 1)	0.67	0.20	0.69	0.23	0.72	0.20	0.68	0.23
Post-test score (out of 1)	0.83	0.10	0.82	0.16	0.86	0.16	0.85	0.17

Table 2. Definition of variables used for the data analyses

Variable	Definition
Prop_Learn_Gain	(post-test score – pre-test score) / (1 – pre-test score), where scores are calculated only over questions relevant to the 2 initial sub-goals
Strategy_Processes	Ratio per period of 10 min. of SRL strategy processes (summary, coordination of information sources, re-reading, note-taking) deployed (agent and user-initiated), normalized over the session time
Monitoring_Processes	Same as above for monitoring processes (feeling of knowing [FOK], judgment of learning [JOL], content evaluation [CE], prior knowledge activation [PKA], monitoring progress to goals [MPTG])
User-init_SRL_first30 (*)	Ratio per period of 10 min of user-initiated SRL processes (monitoring and strategy) during the first 30 min. of the session
User-init_SRL_last30	Same as (*) during the last 30 min.
Agent-init_SRL_first30	Same as (*) for agent-initiated processes
Agent-init_SRL_last30	Same as (*) for agent-initiated processes during the last 30 min.

The experiment involved two sessions: the first one (40 min. long) was used to collect information about participants and for them to take a pre-test on the circulatory system. In the second session (90 min. long), participants activated prior knowledge, set up two sub-goals and then spent 60 min. browsing through the content. At the end,

participants were given a post-test and asked to complete a questionnaire about the PAs. In addition to the variables described in Table 3, we used participants' replies to 2 sets of post-session questions on the quality PAs' feedback (from 1 [very dissatisfied] to 7 [very satisfied]), and the prompts frequency (more, less, or neither more nor less). In FP and FQP conditions, participants were asked if they noticed changes in the prompts frequency (and if yes, if it was an increase, a decrease or irregular variations). Only Mary (monitoring) and Sam (learning strategies) are considered here.

3 Results

Evolution of the Probability of Activation of Rules. The probability of activation of strategy and monitoring rules decreased throughout the session: more in FP than in FQP (as the probabilities could increase), and more for monitoring than for strategy processes. The proportion of user-initiated processes in the probability of activation of processes was lower at the end of the session for learners in FQP than in FP.

Table 3. Summary of Follow-Up ANOVA results and means and std. dev. of variables used

Variable	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2	NP		FP&FQP	
					<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Prop_Learn_Gain	1, 38	2.44	0.127	0.06	40.51	34.53	58.24	37.22
Strategy_Processes	1, 38	18.71	0.00**	0.33	0.12	0.11	0.32	0.20
Monitoring_Processes	1, 38	60.60	0.00**	0.62	0.12	0.06	0.28	0.07

Comparison of SRL Processes Across Conditions. An omnibus MANOVA to compare two conditions (NP vs. FP&FQP) regarding three variables indicated a significant multivariate difference between them for strategy and monitoring processes, Wilk's Lambda = 0.31, $F(3, 36) = 26.78$, $p < .01$, $\eta_p^2 = .69$ (cf. Table 3).

Evolution of the Number of User and Agent-Initiated Prompts. A repeated measures ANOVA revealed a significant main effect of condition on user-initiated SRL behavior, $F(1,38) = 7.64$, $p < .01$, $n_p^2 = 0.17$, but no main effect for time or interaction between condition and time. Another repeated measures ANOVA revealed a significant main effect of time, $F(1,38) = 32.79$, $p < .01$, $n_p^2 = 0.46$, a main effect of condition $F(1,38) = 71.23$, $p < .01$, $n_p^2 = 0.65$, and an interaction of condition and time $F(1,38) = 22.48$, $p < .01$, $n_p^2 = 0.37$, on learners' agent-initiated SRL behavior. Moreover, Table 4 shows that in the FP&FQP conditions, although the number of agent-initiated SRL processes in the last 30 min. is inferior by 40% on average to the one in the first 30 min., the number of user-initiated ones increased (75% of the learners in FP&FQP conditions initiated more SRL processes in the last 30 min. than in the first 30). In the NP condition, learners initiated overall less processes.

Perception of the Agent-Initiated Prompts Frequency Evolution. Overall, participants in FP&FQP did not perceive the decrease in the number of prompts received from Mary and Sam (cf. upper Table 5). Although a majority perceived a frequency

change, as many learners reported an increase as those correctly reporting a decrease (even in FP where, by design, the probability of activation could only decrease).

Satisfaction with PAs' Feedback Quality and Quantity. A one-way ANOVAs revealed a significant difference between conditions for Sam, $F(1,38) = 6.40, p < .05, n_p^2 = .14$, where participants in the NP condition ($M = 4.65, SD = 1.63$) reported higher levels of satisfaction with Sam than those in the FP&FQP conditions ($M = 3.45, SD = 1.36$). No significant difference existed between the NP ($M = 4.60, SD = 1.57$) and FP & FQP ($M = 4.00, SD = 1.92$) for Mary $F(1,38) = 1.17, p > .05, n_p^2 = 0.03$. The lower part of Table 5 shows a majority of participants in conditions FP&FQP would have liked less prompts from Sam, while a majority of participants in condition NP were fine with their frequency. Their opinion about Mary was mixed.

Table 4. Comparison across conditions of user and agent-initiated processes during the session

Variable	NP		FP & FQP	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
User-init_SRL_first30 / _last30	0.90 / 0.95	1.15 / 1.16	1.82 / 2.22	1.60 / 1.48
Agent-init_SRL_first30 / _last30	0.95 / 0.82	0.58 / 0.62	3.47 / 2.05	1.00 / 1.00

Table 5. Proportion of self-reported perception of prompts frequency and satisfaction about it

Perceived frequency of prompts	Mary (monitoring)			Sam (strategy)		
	FP	FQP	FP&FQP	FP	FQP	FP&FQP
Did not change	37.5	16.7	25	25	8.3	15
Decreased	25	33.3	30	37.5	33.3	35
Varied	25	16.7	20	0	33.3	20
Increased	12.5	33.3	25	37.5	25	30
Would have wanted prompts	NP		FP&FQP	NP		FP&FQP
Less frequently	5		40	20		70
More frequently	25		30	25		5
Neither more nor less frequently	70		30	55		25

4 Conclusion and Future Directions

In this paper, we tested the impact of varying the dynamic prompting delivered by MetaTutor's PAs on learners' performance with the system. Preliminary data shows that learners in the FP&FQP conditions enacted consistently more monitoring and strategy SRL processes and had (non sig.) higher proportional learning gains than in the control NP one. The decrease in agent-initiated processes was compensated by a (non sig.) increase in user-initiated ones. Overall, learners did not perceive the prompts variations, but it negatively affected their perception of the quality of the feedback provided by the PAs. Current work focuses on increasing sample size to analyze the impact of the feedback quality on SRL processes enactment.

References

1. Graesser, A.C., Conley, M.W., Olney, A.M.: Intelligent tutoring systems. In: Graham, S., Harris, K. (eds.) *APA Educational Psychology Handbook: Applications to Learning and Teaching*, Washington, DC, vol. 3, pp. 451–473 (2012)
2. Woolf, B.P.: *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Morgan Kaufmann (2008)
3. Azevedo, R., Moos, D.C., Johnson, A.M., Chauncey, A.D.: Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist* 45, 210–223 (2010)