Evaluating Adaptive Pedagogical Agents' Prompting Strategies Effect on Students' Emotions

François Bouchet, Jason M Harley, Roger Azevedo

Evaluating adaptive pedagogical agents' prompting strategies effect on students' emotions

François Bouchet[1] [0000-0001-9436-1250], Jason M. Harley[2] and Roger Azevedo[3]

[1] Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, LIP6, 75005 Paris, France
[2] University of Alberta, Educational Psychology, Edmonton, AB, Canada
[3] North Carolina State University, Psychology, Raleigh, NC, USA
francois.bouchet@lip6.fr, jharley1@ualberta.ca, razeved@ncsu.edu

Abstract. Adapting ITSs that promote the use of metacognitive strategies can sometimes lead to intense prompting, at least initially, to the point that there is a risk of it feeling counterproductive. In this paper, we examine the impact of different prompting strategies on self-reported agent-directed emotions in an ITS that scaffolds students' use of self-regulated learning (SRL) strategies, taking into account students' prior knowledge. Results indicate that more intense initial prompting can indeed lead to increased frustration, and sometimes boredom even toward pedagogical agents that are perceived as competent. When considering prior knowledge, results also show that this strategy induces a significantly different higher level of confusion in low prior knowledge students when compared to high prior knowledge students. This result is consistent with the fact that higher prior knowledge students tend to be better at self-regulating their learning, and it could also indicate that some low prior knowledge students may be on their path to a better understanding of the value of SRL.

Keywords: adaptivity, prompting, pedagogical agents, intelligent tutoring systems, emotions, affects, metacognition, self-regulated learning.

1 Introduction

Adaptation is key to successful learning with intelligent tutoring systems (ITSs), as learners integrate instructional material with elements ITSs are designed to enhance: learning and problem solving [1]. But an efficient short-term teaching strategy can backfire if the learner starts experiencing negative emotions, including those directed toward the tutor. Many studies have shown the benefit of metacognitive prompting to encourage learners to deploy self-regulated learning (SRL) strategies [2]. The idea is to move from co-regulated learning [3], with the assistance of a human tutor or of a pedagogical agent, toward real self-regulation, with the assumption that these skills will transfer beyond the learning session. Because of the relatively short time of interaction with the ITS, one can be tempted to initially choose a high-frequency prompting strategy which progressively decreases as the learner engages in monitoring their learning and deploying learning strategy on their own — a strategy which has been shown to be beneficial for the use of SRL processes, without reducing the perceived

usefulness of the ITS [4]. However, the more insistent or more frequent the interventions, the more likely they may be to trigger emotional reactions from learners. This is important because emotions are critical to fostering motivation and the use of SRL strategies [5]. Therefore, there is a potential risk that eliciting negative emotions through intensive SRL (e.g., cognitive strategies, metacognitive monitoring) process prompting could eventually lead to a form of non-compliance or contempt toward the tutor, and potentially minimize the benefits of SRL. Moreover, as learners with high prior knowledge deploy different SRL strategies from learners with low prior knowledge [6], it is likely their reactions to adaptive prompting could differ.

In this study, we investigated agent-directed emotions self-reported by learners after their interaction with an open-ended learning environment embedded with several pedagogical agents (PAs) that promoted different facets of SRL. We focused particularly on negatively-valenced emotions because of the deleterious impact they can have on learning (in our case, the benefits of SRL) – although exceptions exist such as confusion, which sometimes is positively associated with deep learning [7]. Specifically, we examined three separate but related research questions: (RQ1) Did the frequency of prompts from different PAs affect emotions directed toward them? (RQ2) How did students report feeling toward PAs in two different prompting rule conditions (an initially high but decreasing prompting strategy and a lower intensity prompting strategy)? And (RQ3) whether high prior knowledge students' emotional reactions to adaptive prompting differed from low prior knowledge students? We hypothesized that the higher the number of prompts from an agent, the more likely a learner would be to experience negatively-valenced emotions. We also hypothesized that students with higher initial prompting would be likely to experience more negatively-valanced emotions, and that higher prior knowledge learners would also experience more negative emotions from being told what they may already know with more prompts to self-regulate. We did not investigate the effect of the adaptive prompting on learning outcomes because we have already shown its neutral to mildly positive effects in a previous work [4].

The emotions learners experience while interacting with ITSs is a widely studied topic, and one examined using a variety of methodologies. As with traditional psychological studies, self-report measures are popular instruments for measuring emotions. The Academic Emotions Questionnaire (AEQ) [8] is one such measure; it has been used in research with MetaTutor (but not considered here) and BioWorld [9, 10]. The On-line Motivation Questionnaire (OMQ) [11] has also been used in experiments with iSTART [12] or BioWorld [13] to measure emotions. In addition to self-report measures, ITS researchers have also used various behavioral and physiological approaches, including facial expression recognition (e.g. CERT [14]), electrodermal activity [15], and body language [16]. Online and multimodal approaches to emotion measurement, provide many advantages over self-report, but were not as relevant for our research questions here because we are particularly interested in emotions directed towards a specific part of the ITS: the pedagogical agents. As such, it was most appropriate to use a self-report measure to focus students' answers.

2 Method

2.1 Participants and experimental conditions

One hundred and sixteen undergraduate students (N = 116, 17-31 years old, M = 20.9 years, SD = 2.4; 64.6% female; 62.9% Caucasian) from two North American Universities, studying different majors and with various levels of prior knowledge participated in this study. Each participant received $50 upon completion of the study and was randomly assigned to one of three conditions: (1) *non-adaptive prompt* (NP – n = 58), (2) *frequency-based adaptive prompt* (FP – n = 29) and (3) *frequency and quality-based adaptive prompt* (FQP – n = 29). Participants from the adaptive conditions, FP and FQP, were grouped in some analyses, leading to two samples of identical sizes.

Non-adaptive prompt (NP). In the NP condition, learners received a moderate but constant amount of prompts from the PAs (on average, 1 per 10 minutes) to engage in various SRL processes throughout the learning session. Previous PA prompts, learners' initiative to enact SRL processes, validity of learners' metacognitive judgments or efficiency at using a learning strategy had no impact on the prompts from the PAs.

Frequency-based adaptive prompt (FP). In the FP condition, learners received more prompts at the beginning of the session (on average, 3.5 per 10 minutes), but the probability of both categories of prompts (monitoring and strategy) being triggered decreased after each new prompt was received and after each self-initiated enactment of an SRL process by the learner. Accordingly, participants who had been prompted frequently at first *and* who had been self-initiating SRL processes regularly could potentially end up receiving no further prompts by the end of the session.

Frequency and quality-based adaptive (FQP). The FQP condition applies the same prompt deceasing rules as the FP condition with the addition of two further rules that (if triggered) will *increase* the probability of monitoring or learning strategy prompts of being triggered. If (1) the learner does not comply with a PA's (non-mandatory) prompt to deploy a certain learning strategy (i.e., to re-read a page), or (2) a learner's metacognitive judgment was inaccurate (e.g., selected a page as relevant to his/her active learning when it was not; cf. Table 1 for the list of conditions of success) then the probability of both categories of prompts being triggered will increase.

Table 1. Condition of successes associated to the different type of SRL prompts.

Type	Type of PA's prompt	Condition of success
Monitoring	Judgment of Learning (JOL)	Accurate evaluation of what has been learnt
	Feeling of Knowing (FOK)	Accurate evaluation of what is already known
	Content Evaluation (CE)	Accurate evaluation of the relevance of the content relative to the active sub-goal
	Management of Progress Toward Goal (MPTG)	Learner validates their sub-goal in the next 45s
Strategy	Summarization (SUMM)	If learner delays, must be performed later on
	Coordination of Information Sources (COIS)	Image is opened in the next 45s
	Draw image already opened	Digital notepad in the next 45s
	Draw image not opened yet	Learner accepts to open the image

2.2 The testbed system, experimental procedure and data used

System overview. MetaTutor is an intelligent, hypermedia learning environment in which four embedded PAs help the student to learn more efficiently by prompting them to engage in SRL processes (*cf.* Figure 1). They navigate through the 38 pages (with text and images) on human circulatory system using a table of contents (noted B in Fig. 1). Progress toward the overall learning goal and the sub-goals chosen at the beginning of the session is always visible at the top of the system interface (C in Fig. 1). A timer displays the time remaining in the learning session (A in Fig. 1). One of the four PAs is always visible in the top right-hand corner of the interface (D in Fig. 1), corresponding to the last one who interacted with the student (using text and voice as output, but text-only as input for students' answers to the prompts). The PAs' appearances and voices are the same in each experimental condition, and each PA is comparable in terms of visual and audio quality. Each PA has a specific role: *Pam the Planner* helps the student to plan their learning sub-goals, *Mary the Monitor* helps in monitoring the learning, *Sam the Strategizer* assists with the deployment of learning strategies and *Gavin the Guide* introduces the system and its questionnaires. PAs' prompts are triggered depending on parameters such as the time spent on a page or the relevance of the page to students' current sub-goal. Additional parameters allow to adjust the triggering to obtain an overall higher/lower frequency of prompts (conditions FP and FQP) and to consider compliance and accuracy of previous SRL processes (condition FQP). Below the PA, a palette of buttons allows students to self-initiate SRL processes, leading to a sequence of steps very similar to when the prompt comes from a PA: an invitation to perform the process followed by a feedback on its validity (e.g. agreeing the page is relevant to the current learning sub-goal).

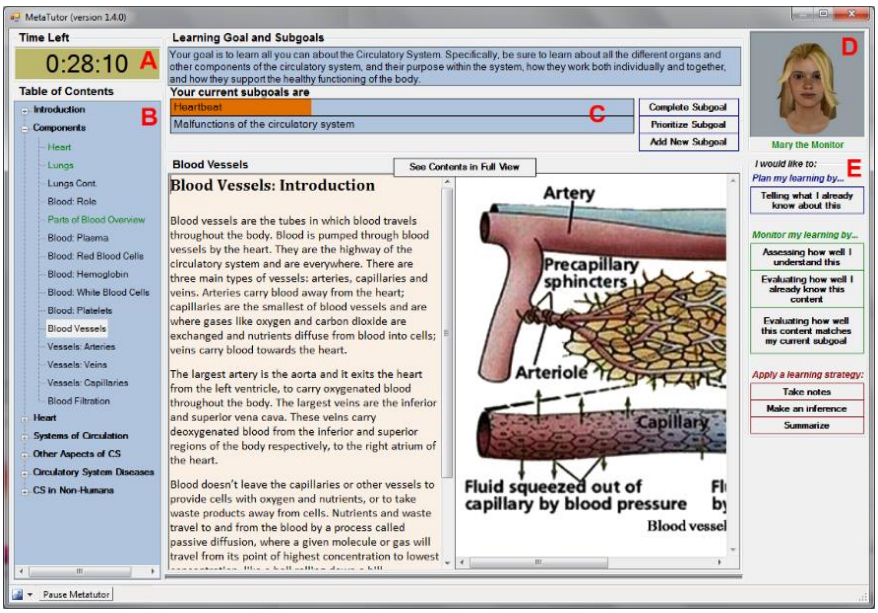

Fig. 1. Annotated screenshot of the system interface.

Experimental procedure. Participants used the system individually on a desktop computer in two sessions separated by one hour to three days. During session 1 (30 to 40 min. long), they filled and signed a consent form and completed several computer-based self-report questionnaires, a demographics survey and a 25-item pre-test on the circulatory system. During session 2 (90 min. long), participants used MetaTutor to learn about the circulatory system. Participants had 60 minutes to interact with the content during which they could initiate SRL processes or do so after a PA's prompt. MetaTutor was paused when participants were watching a video, taking a survey, and during an optional 5 minutes break half-way through the session. At the end of the session, participants were given a post-test and filled a questionnaire, the Agent Response Inventory (ARI) [17], which included statements on the emotions each agent made them feel (e.g. "SAM made me feel frustrated") that they had to rate on a 5-point Likert scale (from "strongly disagree" to "strongly agree").

Data coding and scoring. Because only prompts from Mary and Sam varied between conditions, we focused on emotions toward these two agents. 19 emotions were assessed, but we focused on the negatively-valenced ones (the most deleterious on learning, as mentioned before). When two emotions were very close from each other (e.g. anger/frustration, fear/anxiety, disgust/contempt) we also chose to remove one of the two, on the basis that non-expert students could fail to grasp the real but subtle nuance that exists. In each case, we kept the emotion in the pair that seemed to be the more learning-oriented (e.g. frustration over anger, anxiety over fear) or social (contempt over disgust) one. We ended up with a set of 7 emotions: frustration, anxiety, shame, hopelessness, boredom, contempt, and confusion.

To evaluate the frequency of prompts received by each participant in each condition, we extracted from log-file data the average number of prompts they received from each PA over a period of 10 minutes. Finally, to determine prior knowledge level, we used the adjusted ratio (between 0 and 1) of correct answers in the pre-test[1]. We conducted a median-split on participants' adjusted pretest score, such that participants whose scores fell below the median were labeled as low prior knowledge (LPK) and those who scored above were labeled as high prior knowledge (HPK). 2 participants whose score was equal to the median value (0.727) were excluded. Scores in the LPK group (n = 57) varied from 0.125 to 0.722 (M = 0.523 and SD = 0.149) and scores in the HPK group (n = 57) varied from 0.733 to 1 (M = 0.891, SD = 0.072).

3 Results

3.1 Effect of agents' prompts frequency on agent-directed emotions

Pearson product-moment correlations were run to determine the relationship between number of prompts per period of 10 minutes from Mary/Sam and the score of each emotion toward Mary/Sam. There were significant positive correlations between (a)

[1] We selected only items among the 25 questions that were relative to the subgoals each participant set at the beginning of their learning session (as participants did not have time to explore all the learning material relative to each of the 7 subgoals available)

number of prompts and frustration toward Mary (r = .238, p = .010), (b) frustration toward Sam (r = .338, p = .000), and (c) boredom toward Mary (r = 0.190, p = .041). Other emotions were not statistically significantly correlated.

3.2 Effect of adaptive prompting on agent-directed emotions

Mann-Whitney tests were run to examine differences between learners in conditions NP and FP&FQP in terms of emotions toward Mary/Sam. The results indicated that frustration was higher toward Mary (U = 1155.5, p = .001) in condition FP&FQP (M = 2.83) than in NP (M = 2.10). Regarding Sam, results indicated that frustration was higher (U = 1274, p = .010) in condition FP&FQP (M = 2.66) than in NP (M = 3.31), and that shame was marginally lower (U = 1886.5, p = .091) in FP&FQP (M = 1.47) than in NP (M = 1.81). No statistically significant results were found for the other emotions (*cf.* Table 2).

Table 2. Average self-reported emotions towards Sam and Mary in both conditions

Condition	**NP**				**FP&FQP**			
Agent	**Sam**		**Mary**		**Sam**		**Mary**	
Emotion	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Frustration	2.66**	1.57	2.10***	1.19	3.31**	1.43	2.85***	1.42
Anxiety	2.26	1.40	2.21	1.28	2.17	1.32	2.16	1.32
Shame	1.81*	1.21	1.62	0.93	1.47*	0.90	1.67	1.14
Hopelessness	1.69	1.12	1.55	0.89	1.51	1.02	1.57	1.07
Boredom	2.38	1.22	2.12	1.16	2.53	1.45	2.28	1.34
Contempt	2.31	1.47	1.90	1.13	1.98	1.35	1.91	1.36
Confusion	2.02	1.42	1.60	0.93	1.90	1.31	1.79	1.24

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

3.3 Effect of prior knowledge on agent-directed emotions

First, we examined the existence of an interaction between prior knowledge (groups LPK and HPK) and conditions. We ran multiple two-way ANOVAs to examine the effect of prior knowledge and condition on emotions toward Mary and Sam. No statistically significant interaction between prior knowledge and condition was revealed, leading us to consider prior knowledge individually.

Then as in 3.2, we ran two sets of Mann-Whitney tests to examine differences between HPK and LPK learners in (1) condition NP (*cf.* Table 3), and condition (2; merged) FP&FQP (*cf.* Table 4). In condition NP, no statistically significant results were found between LPK and HPK learners for the 7 emotions tested. In condition FP&FQP, however, the tests revealed that hopelessness was higher toward Sam (U = 499, p = .009) for LPK (M = 1.92) than for HPK learners (M = 1.22). It was also the case with higher confusion (U = 501, p = .016) for LPK (M = 2.38) than for HPK (M = 1.56). Conversely, HPK learners reported marginally more contempt (U = 298.5, p = .064) and frustration (U = 289, p = .051) towards Sam. Similar patterns were found for Mary who elicited more confusion (U = 494, p = .021) for LPK (M = 2.08) than for HPK (M = 1.59), and marginally more hopelessness for LPK (U = 448, p = .097).

Table 3. Average self-reported emotions towards Sam and Mary in condition NP

Condition	**Low prior knowledge**				**High prior knowledge**			
Agent	**Sam**		**Mary**		**Sam**		**Mary**	
Emotion	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Frustration	2.49	1.54	2.06	1.25	2.88	1.58	2.16	1.08
Anxiety	2.27	1.44	2.09	1.31	2.24	1.34	2.36	1.23
Shame	1.79	1.20	1.58	0.89	1.84	1.22	1.68	0.97
Hopelessness	1.73	1.14	1.63	0.98	1.64	1.09	1.44	0.75
Boredom	2.24	1.35	2.27	1.36	2.56	0.98	1.92	0.80
Contempt	2.36	1.47	1.94	1.13	2.24	1.45	1.84	1.12
Confusion	1.91	1.42	1.61	0.95	2.16	1.41	1.60	0.89

Table 4. Average self-reported emotions towards Sam and Mary in condition FP&FQP

Condition	**Low prior knowledge**				**High prior knowledge**			
Agent	**Sam**		**Mary**		**Sam**		**Mary**	
Emotion	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Frustration	3.00*	1.47	2.83	1.31	3.63*	1.34	2.84	1.52
Anxiety	2.08	1.36	2.00	1.19	2.19	1.29	2.22	1.41
Shame	1.63	0.95	1.67	1.07	1.38	0.86	1.69	1.21
Hopelessness	1.92***	1.19	1.71*	1.02	1.22***	0.78	1.41*	1.09
Boredom	2.38	1.32	2.25	1.20	2.63	1.56	2.28	1.46
Contempt	1.71*	1.10	2.00	1.29	2.22*	1.49	1.88	1.43
Confusion	2.38**	1.41	2.08**	1.12	1.56**	1.14	1.59**	1.32

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

4 Discussion

The first two results confirm our initial hypothesis: both agents elicited more negative emotions when more prompts were received, even if their frequency had decreased by the end of the learning session (for most participants, it was below the frequency of prompts received in condition NP). We can assume that the frustration associated with both agents was mostly related to the increased disruptions in the learning task, which was probably stronger initially (in conditions FP&FQP) and hadn't decayed by the end. This could be an issue because frustration can be a useful emotion when temporary and directed toward the learning material, but not necessarily if directed toward a tutor. Regarding other negative emotions, the fact that they are different between Mary (whose additional prompting also was accompanied by increased boredom) and Sam (whose additional prompting led to lower level of shame) indicates that the differences in emotions stems from the agents' roles (cf. section 2.2). Overall, Mary's feedback is more immediately helpful to the student, even when it is negative (e.g. "this page is actually not relevant", "you don't seem to know this content as well as you thought"). Repetitiveness of such feedback simply reveals the limits of what the PA can provide. On the contrary, Sam's feedback can be perceived as more judgmental ("your summary was a little long/short") without necessarily being immediately helpful. Repetitiveness leads to an inhibition of the initial shame learners can have when failing to deploy the suggested strategies — although reduced shame would be

positive for learning, the conjunction with increased frustration makes us assume that learners were probably just becoming more dismissive of Sam's feedback.

The third result goes against our initial hypothesis: not only did high prior knowledge participants not feel more frustration toward the agents, but when there was a difference with low prior knowledge students, they reported feeling less negatively-valenced emotions. The fact that no differences existed in condition NP also means that the differences between LPK and HPK students appeared because of the more intense prompting initially (but was not directly related to the total amount of prompts, as shown by the first analysis). We already know that HPK learners deploy their SRL strategies differently from LPK ones [6], and that HPK learners tend to naturally use more the system prompts to regulate their learning [18]. Therefore, the additional confusion for LPK learners can either mean (a) that they did not perceive the point of agents' prompts (or of SRL altogether), and that the increased intensity only made them wonder more about their interest, or (b) that the confusion was only felt initially, which can be a desirable initial state for learning, and that later on in the session, they were starting to perceive the value of agents' prompts. The fact that the system usefulness was not perceived lower in conditions FP&FQP tends to indicate that at least some LPK learners were in that situation [4]. This is an encouraging result, as it shows that despite the limits of the PAs' range of prompts and the repetitiveness of some of their feedback, some low prior knowledge students (who are the ones who can benefit the most from self-regulating their learning) managed to perceive more the value of the self-regulation fostered by the system.

5 Conclusion, limits and future works

Overall, this study shows that adaptive prompting with a more intense initial strategy is a double-edged sword: on the one hand, it emphasizes the limits of the ITS and its embedded PAs, whose smallest flaws become magnified and prone to increased frustration. On the other hand, although high prior knowledge participants quickly seem to understand the benefits of PAs' help (even if it is of low intensity), having more frequent prompts seems to help low prior knowledge participants more, after what we can assume to be a temporary initial confusion. This help is limited, however, by the usefulness of the prompts, which we have previously seen, may not be specific enough to lead to a significantly measurable increase in the learning outcome (when comparing pre to post-test results) [4].

One of the limits of this study is that we only measured participants' feelings toward each agent for the overall session at its end (and at one point in time). We therefore cannot, for example, rule out that LPK students finished their learning session more confused than in the non-adaptive prompting condition. Using constant emotion monitoring, for instance, through automatic facial analysis, could help with this issue [19]. It would also decrease the reliance on self-report, as students can sometimes be poor reporters of their own emotional state. It is worth noting, however, that even if we have previously found good agreement between automatic facial expression analysis and self-report; facial expression recognition software has typically focused on

basic emotions [15], to the exclusion of some achievement-related emotions such as boredom and confusion. Another limit is the fact that we considered conditions FP and FQP together in all analyses on account of sample size. Evaluating these two adaptive conditions separately represents an important future direction Finally, we lack long-term evaluations of whether participants have indeed internalized more the benefit of using externally-prompted, and sometimes collaborative [20], self-regulated learning strategies, which is the goal of a system such as MetaTutor. Using information from students' emotional responses toward MetaTutor and its PAs to provide real-time, user-adaptive [21] SRL prompts also represents an important future direction for this work.

Acknowledgements. Research supported by funding from NSF (DRL 1008282, DRL1431552, DRL 1660878), SSHRC, and CRC program awarded to third author.

References

1. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis. J. Educ. Psychol. 106, 901–918 (2014).
2. Bannert, M., Mengelkamp, C.: Scaffolding Hypermedia Learning Through Metacognitive Prompts. In: Azevedo, R. and Aleven, V. (eds.) International Handbook of Metacognition and Learning Technologies. pp. 171–186. Springer New York (2013).
3. Järvelä, S., Hadwin, A.F.: New Frontiers: Regulating Learning in CSCL. Educ. Psychol. 48, 25–39 (2013).
4. Bouchet, F., Harley, J.M., Azevedo, R.: Can adaptive pedagogical agents' prompting strategies improve students' learning and self-regulation? In: Micarelli, A., Stamper, J., and Panourgia, K. (eds.) Intelligent Tutoring Systems: 13th International Conference. pp. 368–374. Springer International Publishing, Zagreb, Croatia (2016).
5. Pekrun, R., Perry, R.P.: Control-value theory of achievement emotions. In: Pekrun, R. and Linnenbrink-Garcia, L. (eds.) International Handbook of Emotions in Education. pp. 120–141. Routledge (2014).
6. Taub, M., Azevedo, R., Bouchet, F., Khosravifar, B.: Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? Comput. Hum. Behav. 39, 356–367 (2014).
7. D'Mello, S., Lehman, B., Pekrun, R., Graesser, A.: Confusion can be beneficial for learning. Learn. Instr. 29, 153–170 (2014).
8. Pekrun, R., Goetz, T., Titz, W., Perry, R.P.: Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. Educ. Psychol. 37, 91–105 (2002).
9. Harley, J.M., Bouchet, F., Azevedo, R.: Examining How Students' Typical Studying Emotions Relate to Those Experienced While Studying with an ITS. In: Intelligent Tutoring Systems: 14th International Conference. Springer International Publishing, Montreal, QC, Canada (2018).

10. Jarrell, A., Harley, J.M., Lajoie, S., Naismith, L.: Success, failure and emotions: examining the relationship between performance feedback and emotions in diagnostic reasoning. Educ. Technol. Res. Dev. 65, 1263–1284 (2017).
11. Boekaerts, M.: The on-line motivation questionnaire: A self-report instrument to assess students' context sensitivity. In: Pintrich, P.R. and Maehr, M.M. (eds.) New Directions in Measures and Methods. pp. 77–120. Emerald Group Publishing Limited (2002).
12. Jacovina, M.E., Tanner Jackson, G., Snow, E.L., McNamara, D.S.: Timing Game-Based Practice in a Reading Comprehension Strategy Tutor. In: Micarelli, A., Stamper, J., and Panourgia, K. (eds.) Intelligent Tutoring Systems. pp. 59–68. Springer International Publishing (2016).
13. Lajoie, S.P., Naismith, L., Poitras, E., Hong, Y.-J., Cruz-Panesso, I., Ranellucci, J., Mamane, S., Wiseman, J.: Technology-Rich Tools to Support Self-Regulated Learning and Performance in Medicine. In: Azevedo, R. and Aleven, V. (eds.) International Handbook of Metacognition and Learning Technologies. pp. 229–242. Springer New York, New York, NY (2013).
14. Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., Bartlett, M.: The computer expression recognition toolbox (CERT). In: Face and Gesture 2011. pp. 298–305 (2011).
15. Harley, J.M., Bouchet, F., Hussain, M.S., Azevedo, R., Calvo, R.A.: A Multi-Componential Analysis of Emotions during Complex Learning with an Intelligent Multi-agent System. Comput. Hum. Behav. 48, 615–625 (2015).
16. D'Mello, S.K., Graesser, A.: Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. User Model. User-Adapt. Interact. 20, 147–187 (2010).
17. Harley, J.M., Carter, C.K., Papaioannou, N., Bouchet, F., Landis, R.S., Azevedo, R., Karabachian, L.: Examining the predictive relationship between personality and emotion traits and students' agent-directed emotions: towards emotionally-adaptive agent-based learning environments. User Model. User-Adapt. Interact. 26, 177–219 (2016).
18. Bouchet, F., Kinnebrew, J.S., Biswas, G., Azevedo, R.: Identifying Students' Characteristic Learning Behaviors in an Intelligent Tutoring System Fostering Self-Regulated Learning. In: Yacef, K., Zaïane, O., Hershkovitz, A., Yudelson, M., and Stamper, J. (eds.) Proceedings of the 5th International Conference on Educational Data Mining. pp. 65–72. , Chania, Greece (2012).
19. Mudrick, N., Rowe, J.P., Taub, M., Lester, J.C., Azevedo, R.: Toward affect-sensitive virtual human tutors: The influence of facial expressions on learning and emotion. In: Busso, C. and Epps, J. (eds.) Proceedings of the 2017 seventh international conference on affective compu-ting and intelligent interaction (ACII). pp. 184–189. IEEE Computer Society, Washington, DC, USA (2017).
20. Harley, J.M., Taub, M., Azevedo, R., Bouchet, F.: "Let's set up some subgoals": Understanding human-pedagogical agent collaborations and their implications for learning and prompt and feedback compliance. IEEE Trans. Learn. Technol. 11, 54–66 (2018).
21. Harley, J.M., Lajoie, S.P., Frasson, C., Hall, N.C.: Developing Emotion-Aware, Advanced Learning Technologies: A Taxonomy of Approaches and Features. Int. J. Artif. Intell. Educ. 27, 268–297 (2017).