Off-task Engagement in a Teachable Agent based Math Game

Betty TÄRNING[a*], Magnus HAAKE[b] & Agneta GULZ[a]
[a]*Lund University Cognitive Science,* [b]*Department of Design Sciences, Lund University*
*betty.tarning@lucs.lu.se

Abstract: A previous study compared two student groups that played a mathematics game based on a teachable agent. One group played with, and the other without, the inclusion of a social conversation module: a chat between the student and the teachable agent. Results were that students who used the game with the chat included had a more positive experience of the game and learned more in the sense of teaching their agent better. However, patterns differed between sub-groups of students. Low-achievers did not prefer the game with the chat included, whereas high- and mid-achievers did, but simultaneously low-achievers tended to chat more. Low-achievers tended not to use the options of not starting the chat or quitting a chat beforehand as much as high- and mid-achievers did. In this paper we pursue a more detailed analysis of the students' conversational behavior in the chat. The analytic focus is on the notion of engagement. Results point towards differences between the student groups in their engagement in the off-task conversation, that in turn can help explain the previous somewhat paradoxical result.

Keywords: teachable agent, off-task conversation, natural language dialogue, engagement, low- and high-achievers, quality of conversation, learning

Introduction

The paper approaches engagement in off-task conversation between students and a teachable agent. The starting point was an intriguing result from a previous study regarding how math low- and high-achievers respectively responded towards off-task conversation. We wanted to reach a better understanding of this result by undertaking an additional analysis. In this the notion of *engagement* became the analytic focus.

1. Background: the original result of different patterns for low- and high achievers

The underlying mathematics game trains basic arithmetic skills [1,2], and the student teaches her *Teachable Agent*, TA, to play the game. For more details we refer to [2]. The focus in this paper is on the chat-like conversation between student and TA by means of natural language text input. We refer to this as *off-task conversation* as opposed to the *on-task conversation* between student and TA, which is a multiple-choice guided conversation during the math game sessions, targeted at mathematical content. The motive behind the chat is to enhance students' experiences and increase their inclination to use the game over time. Yet another motive is to enable additional pedagogical interventions such as influencing students' math self-efficacy, cf. [3]. Three classes in a Swedish school participated in the original study, with 38 female and 42 male 12-14 year olds. 18 were by their teachers classified as math low-achievers, 39 as math mid-achievers and 23 as math high-achievers. Each class was divided into two groups with an even distribution with respect to

math achievement and gender in each group. All students got to play the game during three lessons. The *WithC*-group used the game with, and the *NoC*-group without the chat module. For switching between playing the game and chatting, the TA uses the metaphor of "breaks" between lessons in school. During the first three "breaks" the chatting with the agent was mandatory until the "break" ended after three minutes. After this students were always offered the *choice* between taking a "break" to chat with the agent or continuing to play the game. They could also choose to end the chat before the "break" was over. However students in both conditions did get breaks in the sense of cognitive rest and change of activity. Instead of training their TA, they could let it play against the computer which only requires passive viewing. It was obvious that students did make use of the latter as a kind of break.

After having completed the third lesson, students filled out a questionnaire regarding their experience of the game. In sum the results showed (for details see [1]) that students in the *WithC*-group tended to have a more positive game experience than students in the *NoC*-group (diff=0.54, p=0.07). Furthermore, students in the *WithC*-group tended to reach better result in terms of how well they taught their TA (diff=0,3, p=0.07). But separating the results into student sub-groups, revealed that it was only the medium- and high-achieving students that reported a more positive experience of the game for the *WithC* compared to the *NoC* condition. The *low-achievers* did not seem to prefer any of the two system versions before the other. A comparison of the tendency to refrain from starting a chat or quitting a chat before chat time was out also revealed different patterns in the subgroups. High-achievers, and to some extent mid-achievers refrained from starting a chat or quitting a chat beforehand – to a considerably higher extent than low-achievers.

a) **Student:** hola senorita
TA: What do you do on the weekends?
Student: Play handball and just chilling with my friends
TA: OK
Student: and you?
TA: I quite often listen to music
Student: OK, what do you like listening to?
TA: Mostly rock and pop
Student: What's so great about rock do you think?
TA: I did not understand, what do you mean?
Student: Why do you like rock?

b) **Student:** are you good at school?
TA: pretty good
Student: neeerd
TA: can't we talk about spare time interests instead
Student: ok, I like football a lot
TA: how, what do you mean?
Student: It's fun to go for the footballs and hit them
TA: Well, it's difficult for me to understand what you say, can we talk of something else?
Student: Do you have problems with the Swedish language or what?
TA: I'll go to the cinema this weekend, it's so nice just forget about everything and go into another world kind of
Student: Can I come along?
TA: What do you do usually on the weekends?

c) **TA:** maybe we could talk about music
Student: .
TA: I did not understand, what do you mean by that?
Student: .
TA: It's difficult for me to understand what you say, can we talk of something else?
Student: .
TA: I think Coldplay is really cool, which is your favorite artist?
Student: .
TA: well
Student: .
TA: okay
Student: .

d) **Student:** hi
TA: hi
Student: do you like animals?
TA: I don't understand, ask me something else
Student: dfjhasjk
TA: Maybe we can talk about hobbies?
Student: fghjkl
TA: What do you mean?
Student: hfhdfdjdjdj
TA: Yesterday I scored twice in floor ball in gym class, do you like sports?
Student: djdjdjd
TA: OK
Student: djdjd

Figure 1: Examples of chats with high (a & b) and low (c & d) engagement. (Translated from Swedish.)

Taken together there is a seeming paradox in that high-achievers and mid-achievers tended to like the game better when the chat was included (compared to low-achievers) while they also were more inclined to quit and to refrain from starting the chat when they had the option to. In other words: high- and mid-achievers *liked* the software more when there was a chat included, but they *chatted less* than the low-achievers. Low-achievers on the other hand

chatted more but did *not like* the software to any higher degree with the chat included. We therefore decided to conduct a more detailed analysis of the participating students' chat behavior. At first, two researchers did an independent screening of the chat logs in order to familiarize them selves with the material. They made the common observation that the *engagement* in the conversation, as indicated by the chat logs, varied considerably between chats of each individual student as well as between the individual students. Figure 1a & 1b show two examples of parts of chat logs that, according to the researchers coding scheme, indicate high engagement, and Figure 1c & 1d present two examples of parts of chat logs that indicate low engagement.

Engagement is increasingly discussed within the educational technology field, e.g. [4]. The research questions that we hoped to illuminate by the analyses to be presented in this paper were the following: 1) To what degree did the students seem *engaged* in the off-task conversations with their teachable agent? Did low- and high-achievers differ in this respect? 2) What did students do when the chat logs indicate that they are or have become very disengaged in the chat conversation? In particular: did they quit or not? Did they refrain from starting the next chat or not? Did they continue chatting in a similar way indicating low engagement, or not? Did low- and high-achievers differ in this respect?

2. Method

6 = the student is driving the conversation and asks relevant questions, sometimes in the form of a counter question when having answered a question from the TA *("do you have siblings?"; "I feel good right now, how about you?"*

5 = the student is driving the conversation but asks more curious and exploratory questions *("do you have problems to understand the Swedish language?"; "are you gay?")*

4 = the TA urges the student to ask a question and the student gives a relevant answer *(TA: "what type of music do you like?" Student: "techno"; TA: "Ask me something else" Student: "do you like horses?")*

3 = the TA urges the student to ask question and the student gives an answer that is (semi)-relevant but questionable in terms of being engaged in driving the conversation further: *(TA "I don't get it, can we talk about something else?" Student: "You are stupid"); (TA: "it seemed fun to me" student: "ha ha, you nerd").*

2 = the student drives the conversation but asks irrelevant questions *("are you fat?")*

1 = the TA either asks a question or urges the student to ask a question but utterances provided by the student are irrelevant/not furthering the conversation *(TA: "What do you mean by that?" Student: "go and drown yourself"; TA:" I don't get it can we talk about something else?" Student: "asshole")*

0 = nonsense syllables and blank space *(Student: ". "; Student: "xasdfghhhh")*

Figure 2: The coding scheme for the chats with some example utterances. (Translated from Swedish.)

Measuring off-task conversation engagement: Chat logs were collected in the *WithC*-group for 30 students: 11 high-achievers, 13 mid-achievers and 6 low-achievers. Each student was involved in 3 to 8 chats and on the average each student exchanged 130 phrases with their TA. Half of them were uttered by the agent and half by the student. Each phrase produced by a student was categorized in terms of the degree of engagement in the conversation and was given a value between 0 and 6 (0 representing extremely low engagement and 6 very strong engagement, see Figure 2). The context of the conversation so far and the utterance just made by the TA was taken into consideration, with the main objective to estimate to what degree the student's utterance was an engaging conversation initiative, or a suitable and engaged response, given the TA:s previous utterance. Two researchers coded the dialogue (inter-rater reliability measure (Cohen's kappa): $\kappa = 0.86$) and means were calculated.

3. Results and analysis

Engagement in off-task conversation: We chose to calculate two engagement values, one for the first three mandatory chats (*chat 1-3*), and one for the following chats (*chats 4-x*), where the chatting was optional and could be quitted beforehand. What then comes forth is the following (Table 1). For chat 1-3 there is no significant difference in engagement between high- and low-achievers (*t*-test [one-tailed]: $p = 0,171$, but for chat 4 and onwards the engagement value is significantly higher for high achievers than for low achievers (*t*-test [one-tailed]: $p = 0,044$; all participants with 4 or more chat sessions included).

Table 1: Off-task engagement for mandatory chats (chat 1-3); Off-task engagement for optional chats (chat 4-x); all participants with 4 or more chat sessions included.

Groups	off-task (chat 1-3)			off-task (chat 4-x)			one-tailed *t*-tests
	n	Mean	Var.	*n*	Mean	Var.	
low	6	3,32	2,3141	4	2,29	1,2542	chat 1-3: $p_{high,low} = 0,171$
mid	9	4,04	0,6380	11	3,97	1,3725	chat 4-x: $p_{high,low} = 0,044$
high	8	3,93	0,5908	8	3,82	1,9366	

What did students do when disengaged in the chat conversation? This analysis started by an identification of those chat passages where a student clearly seemed to have lost engagement in the conversation with the agent. One criterion is when a student repeats a blank, a dot, a single word, or meaningless strings, and continues to do so without getting back to a productive conversation. Another criterion is when a student goes on with something that seems more like a monologue, sometimes including harassment, which does not relate to any of the utterances by the TA and, again, does not get back to a productive conversation. All chat logs for all participants were evaluated. Fourty-three instances of disengagement according to the criteria above were identified (see Table 2). Out of these, 22 were within high-achiever chats, 11 within low-achiever chats, and 10 within mid-achiever chats. Since there were 11 high-achievers and 6 low-achievers in the *WithC*-group, it was equally common that a low-achiever and a high-achiever did get strongly disengaged. What differed between the two student groups, however, was the behavioral pattern in this kind of situation – even though we cannot claim statistical significance given the limited number of students involved in the analysis.

Table 2: Actions taken in situations where the chat log indicates strong disengagement in the conversation on the part of the student. (All participants and chat logs included.)

	low-achievers (*n* = 6)	mid-achievers (*n* = 13)	high-achievers (*n* = 11)	Σ (sum)
Quits the chat	0	3	**5**	*8*
Does not start next chat	3	2	**13**	*18*
Gets on with unengaged chat	**8**	5	4	*17*
Σ (sum)	*11*	*10*	*22*	*43*

For the 22 instances of low engagement in conversation between high-achievers and the TA, the student quitted the chat 5 times, and refrained from starting the next chat 13 times. Only in four of the instances did the student both continue the disengaged chat and also start the next one. Reversely for the 11 instances with low-achievers, the student continued the disengaged chat and also started the next chat 8 times, with only 3 instances of refraining from starting the next chat and no case of a student quitting a chat.

In other words, low-achievers were more inclined to continue a chat even when there is a strong indication that they were unengaged, whereas high-achievers were more inclined in these situations to take control or action: they quit the chat and/or refrained from starting

next chat. This result is in concordance with the results reported in table 3: the engagement value for chat 4 and onwards, where the student could control the starting and ending of a chat, was significantly higher for high-achievers than for low-achievers.

Two aspects of the differing behavioral patterns may contribute to an explanation of why high-achievers tended to like the game better with the chat included whereas low-achievers did not. First, it is well known that the experience of having control over one's situation plays an important role for a positive learning experience. Second, with the behavioral differences described, low-achievers tended to spend more time than high-achievers with a chat they might have experienced as non-engaging (boring, meaningless).

4. Discussion and conclusion

The analysis presented in this paper did provide us with a possible explanation of the previous somewhat intriguing result that high–achievers but not low-achievers *liked* the software more when there was a chat included, but that they *chatted less* than the low-achievers. The suggested explanation is the different actions taken when getting into a clearly disengaged conversation with the TA. Low-achievers in the study did not tend to take control over a situation of disengagement in the sense of quitting the chat and/or refraining from chatting next time, whereas the high-achievers tended to do so. This can be important, in more general terms, to consider for designers of pedagogical games: how are possibilities of taking control in the game presented and to what extent will different students use these possibilities? For our case this is important since we are aiming for pedagogical interventions regarding math self-efficacy beliefs via the chat. In turn this is most important for low-achievers, and therefore we need a chat that works well for them.

We plan to make more information from the math game sessions accessible to the chat module, so that more detailed conversations about the game play can be conducted in the chat. This plan gets support from the chat-logs from the present study. More than a third of the students spontaneously initiated chat conversations about the math game with their TA, wanting to discuss whether the TA found it difficult, whether the TA thought it had learnt much, etc. Notably this applies to both high- mid- and low-achievers and thus seems a promising venue for pedagogical interventions.

References

[1] Gulz, A., Haake., M., & Silvervarg, A. (2011). Extending a Teachable Agent with a Social Conversation Module – Effects on Student Experiences and Learning. In G. Biswas et al. (Eds.): AIED 2011, LNAI 6738, pp. 106-114, 2011. Springer.
[2] Pareto, L., Arvemo, T., Dahl, Y., Haake., M., & Gulz, A. (2011). A Teachable-Agent Arithmetic Game's effects. In G. Biswas et al. (Eds.): AIED 2011, LNAI 6738, pp. 247-255, 2011. Springer.
[3] Kim, Y., Wei, Q., Xu, B., Ko, Y., & Ilieva, V. (2007). MathGirls: Increasing girls' positive attitudes and self-efficacy through pedagogical agents. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proc. of AIED 2007* (pp. 119-126). Amsterdam, The Netherlands: IOS Press.
[4] Quinn, C. (2005). *Engaging Learning – Designing e-Learning Simulation Games*. San Francisco, CA: John Wiley & Sons, Inc.