Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

I will help you, but will you help me? How the Perception of a Teachable Agent May Influence Performance

Eva-Maria TERNBLAD [a*] Magnus HAAKE [a] & Betty TÄRNING [a]
[a] *Dept of Cognitive Science, Lund University, Sweden*
* eva-maria.ternblad@lucs.lu.se

Abstract: Learning by teaching someone else has proven to be beneficial in both human-human and human-agent interaction. Instructing someone else, that is, taking the role as a tutor, has a series of positive effects on students' learning and performance. For example, the fact that someone else is in need of help seems to affect students to put more effort into the task at hand, trying harder, being more thorough and persisting longer. This has inspired researchers to design educational software that uses teachable agents (TAs). The use of such software has shown to improve students' – not the least lower-achieving students' – learning. Designing teachable agents is, however, a delicate matter since the personality and capability of the agent may affect students' behaviors and performance. This study, which includes data from 156 6th grade students who used an educational game in history during three lessons, contributes to research on how students' perceptions of a neutral TA (without any specific personality and behavior) may influence their performance. The focus is on whether and how students' learning gains are influenced by the extent to which students perceive their TA as someone who does need their help. In addition, we explored whether such potential effects would differ between lower- and higher-achieving students. Results were that students' perceptions about the TA's need for help was a significant predictor of high performance (based on in-game performance and post-test scores) – independent of their general achievement level (grounded in the students' reading proficiency). In other words, all students, whether lower- or higher-achieving, benefitted from being convinced of the agent's need for help. This finding is somewhat different from previous studies, where TAs mainly have been found to be beneficial for lower-achieving students. In relation to this, the present study adds a novel piece of information by suggesting that for TAs to be beneficial in educational software, their need for help should be clearly communicated and emphasized. This, on the other hand, may make them useful for all students, independently of how well they usually perform in school.

Keywords: Teachable agents, protégé effect, educational games, learning by teaching, achievement level

1. Introduction

Pedagogical agents are thriving in a broad range of different pedagogical contexts, in roles as experts, motivators, or mentors (e.g., see Kim & Baylor, 2016). In this paper we will focus on pedagogical agents that themselves can learn, referred to as *teachable agents* or *TAs* (Chase, Chin, Oppezzo & Schwartz, 2009). In the context of digital educational applications, a TA exploits the pedagogical tradition of 'learning by teaching' in that it takes the role as a tutee and learns from the student, while the student him- or herself learns from instructing the TA. Benefits from educational software using TAs have been shown in a number of studies (Biswas, Leelawong, Schwartz, Vye, & The Teachable Agents Group at Vanderbilt, 2005; Chase et al., 2009; Okita & Schwartz, 2013; Sjöden & Gulz, 2015; Tärning & Silvervarg, 2019). One of several documented benefits is the so called *protégé effect* (Chase et al, 2009), i.e., that students put more effort into their own learning (measured as spending more time on tasks, preparing in more detail, etc.) when they learn in order to instruct someone else, compared to when they learn for themselves. As of today, the large bulk of studies on TAs targets learning benefits from using different types of TAs. In contrast, the aim of this study is to investigate how a more neutral

TA is perceived (as a sibling, a classmate, or as an assistant) and whether a student's perception of the TA as someone in much or little need of help affects performance and learning.

1.1 Design features of teachable agents

Even though a TA's role is always that of a tutee, the character can be portrayed with different personalities, and several prior studies have examined the effectiveness of certain agent design features on specific learning outcomes. Ceha, Lee, Nilsen, Goh and Law (2021) for example found that a TA that tells jokes and funny stories for the amusement of others can increase motivation and effort. Further, Silvervarg and Månsson (2018) found that the way a teachable agent is introduced can affect how it is perceived, which, in turn, can influence students' perseverance. In their study, two groups of students interacted with a TA in an educational game. One group only received the introduction to the agent that was built into the software, whereas the other group also received an additional verbal introduction from the research leaders. Students in the latter group were more inclined to perceive the agent as someone with a desire to learn. They also gave higher ratings in self-reports on their experiences of putting in an effort and not giving up. Other studies by Tärning, Gulz, and Haake (2017) and Tärning, Silvervarg, Gulz, and Haake (2019) have shown that a TA portrayed as having low self-efficacy in a specific domain can increase students' performance and self-efficacy in that domain, and especially so for students that themselves have low self-efficacy. The authors here propose that a TA with low self-efficacy might come across as someone in more need of help than a TA with high self-efficacy.

In a similar study, Silvervarg, Kirkegaard, Nirme, Haake and Gulz (2014) investigated how a challenging TA was perceived among students with higher and lower self-efficacy (measured by self-reports on the students' teaching abilities). The researchers found that, in general, students were inclined to go along with the TAs challenging proposals rather than holding on to their own initial response, both when their initial response was correct but to a higher extent when it was incorrect. They also found that students with lower self-efficacy experienced that the TA challenged them too often. This was not the case for students with higher self-efficacy.

Taken together, there is evidence that the way the TA is portrayed can have an impact on how students perceive it, and this also has a potential to affect their learning.

1.2 Helping a teachable agent

As mentioned above, several studies have investigated possible effects of TAs with respect to different roles or different personalities. For instance, Tärning and Silvervarg (2019), analyzed chat dialogues between students and their TA in an educational game, and found that TAs that expressed low self-efficacy with respect to their own ability to learn math received more positive and encouraging comments about their intelligence and competence. They also found that the students responded more frequently to the feedback from such a TA, as compared to a TA that expressed high self-efficacy. Similar to Tärning et al. (2017; 2019) the authors here argue that a TA with low self-efficacy might appear as someone in more need of help and therefore receives a better treatment compared to a TA with high self-efficacy (which might be perceived as someone that can take care of itself, without help).

Furthermore, Ogan et al. (2012) explored students' interaction with a TA (Stacy) in a math game (using the SimStudent platform) in terms of social or non-social conversation. Results were that the students who treated Stacy as a peer and for example spoke directly to her, were more successful in their learning task. This in contrast to students who talked to Stacy in a third person perspective and made fewer social comments, which was correlated with negative learning gains. Evidently, it appears that *how a student perceives their TA* can influence *how they treat it* and that a better treatment, for example in terms of being more polite, can be beneficial for the students' own learning.

Strong positive learning benefits from using TAs have so far been most pronounced in lower-performing students (Chase et al., 2009; Sjöden & Gulz, 2015; Pareto et al., 2009, Tärning et al., 2017; Tärning et al., 2019), and a number of explanatory mechanisms for this have been proposed. One possible explanation is that the TA always is inferior and more ignorant than the student and hence, the student has the opportunity to take the lead as someone more knowledgeable – a situation lower-performing students may not be familiar with in a school context. Seemingly, taking the lead as someone

competent, able to teach someone else, can improve student's self-efficacy. Hence, experiencing that someone else is in need of your help can be a potential incitement for learning.

However, students' individual opinions about a TA may also play a role. The present study therefore addresses possible variations in students' perception of a more neutral TA (i.e., a TA not designed to have any distinctive personality or behavior) and to which extent students perceive the TA as someone in need of help. Furthermore, we also investigate whether this, in turn, has an influence on their performance.

1.3 Research questions

Although the benefits and effects of TAs have been evaluated in several studies, students' attitudes towards their TA and the relation between these and students' learning and performance has been less studied. It has been shown that TAs are beneficial for lower-performing students *as a group* and likewise for students with lower self-efficacy *as a group*, but we know less about individual variability *within* such groups. The way a student perceives her agent may also tell us something about the student's self-awareness of her role as a teacher. Therefore, we are not only interested in students' characterization of their agent as such (for example as a sibling or a classmate), but also in their ideas about the agent's need for help. To our knowledge, no previous study has explored these issues. Consequently, we end up with the following research questions (RQs):

RQ.1: How do students characterize their (teachable) agent (when they can choose among a sibling, a classmate, and an assistant) and to what extent do they think that their agent is in need of help?

RQ.2: Will students with different achievement levels characterize their teachable agent differently?

RQ.3: Will student's opinion of their teachable agent's need of help influence their performance, and will such a potential impact differ between higher- and lower-achieving students?

2. Method

The research data in this study originates from a larger study on feedback engagement performed in spring 2019. A presentation of the participants and the experimental setup, together with a description of the material and a definition of the parameters relevant for this particular study, follows below.

2.1 The educational game

The stimulus in the study consists of an educational game in history where the students make time-travels to historical scenes and scientists, search for text-based information, and solve tasks (constructing a concept map, setting up a timeline, sorting statements into categories, or answering multiple-choice questions). To be able to continue and progress in the game, the tasks (six in total), must be completed one at a time with a sufficient degree of correctness (80%). The students have, however, unlimited attempts to revise the tasks, and they are allowed to make as many time-travels as they want.

The game contains two agents: Professor Chronos (a middle-aged wizard-like gentleman) and the time-elf Timy (an androgyn childlike character). Professor Chronos, who is the current 'Guardian of History', is about to retire and needs a successor. The educational narrative can then be set to one of two conditions. (i) *Learning for oneself;* The student's mission is to become Professor Chronos' successor and hence needs to prove his or her proficiency in history by making time-travels and solving tasks. In this condition, Timy has a role of a narrator. (ii) *Learning by teaching*; The time-elf Timy wants to become Professor Chronos' successor, but since he/she is suffering from time-travel sickness, the student has to make the time-travels and instruct Timy on how to solve the tasks. Timy then presents the solutions to Professor Chronos. In this condition, Timy has the role of a TA.

In the first condition, Timy in the role of a narrator has a less prominent role compared to the second condition where Timy has the role of a TA (Figure 1). In the first condition, Timy welcomes the

student to the game world and presents the game narrative without engaging in any dialogue with the student. After that, Timy's role in the first condition is restricted to introducing the tasks and instructing the student on how to solve them, e.g., dragging items to boxes or putting figures on a timeline. In the TA-condition, on the contrary, Timy also interacts with the student throughout the game by means of shorter dialogues. Figure 2 below shows Timy introducing the game (left) and having a dialogue with the student in his/her role as TA (right).

Timy as a narrator	Timy the time-elf introduces the game. No dialogues.	Timy presents the tasks.	The student presents the solutions of the tasks to Professor Chronos without Timy's presence. No dialogues.
Timy as a TA	*Timy the time-elf introduces the game as well as her/himself as a teachable agent. A shorter dialogue between the student and Timy.*	*Timy presents the tasks.*	*Timy is present during the tasks where the student instructs him/her. Timy then presents the solution to Professor Chronos. Some shorter dialogues between the student and Timy.*

Figure 1. Narrative of the game and the role of the time-elf in the two versions of the game – representing the two conditions in the study.

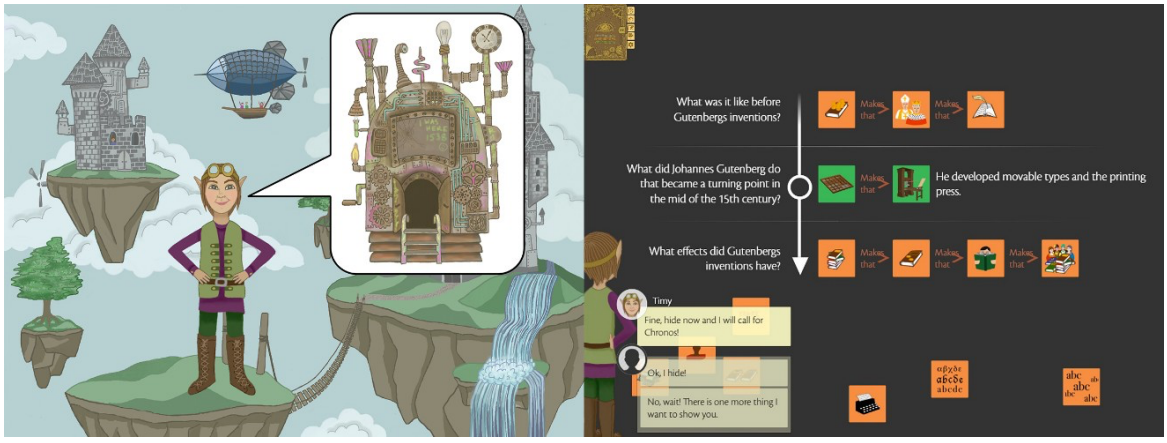

Figure 2. Left: Timy introducing the game narrative (both conditions); right: Timy having a dialogue with the student when being instructed on a task (TA-condition).

2.2 Participants, materials and procedure

In the original study, 285 Swedish 6th grade students in eleven classes from six different schools in southern Sweden participated (see Silvervarg, Wolf, Blair, Haake & Gulz, 2021). But, due to schedule difficulties and shortage of time, only 156 students were able to respond to our 'agent opinion' questionnaire (described below) and report on their personal experience of the game and how they perceived their TA. This data makes up the basis for this study.

The participants played the game during 3 separate sessions, each lasting approximately 60 minutes. During each session, two experimenters were on site, helping the students with questions and technical issues regarding the game. All students were randomly distributed to one of the two conditions ('Timy as TA' or 'Timy as narrator'), and each class was therefore divided into two groups (one for each condition), working in separate classrooms. At the first session, the participants were introduced to the game through a short film with Timy presenting the game narrative (see Figure 2, left).

Since the ability to process text-based information has a strong impact on a student's overall performance in the game, the teachers provided assessments (low, mid, high) of each student's reading proficiency. However, when observing the students solving their first task, it became clear for the researchers that the ability for the 'mid-achievers' to make use of the information in the game were quite diversified (since some of them solved the first task – a simple sorting task – quite easily, while

others failed repeatedly). This heterogeneity was not perceived within the other two groups. As a consequence, based on their initial in-game performance (more precisely: the number of revisions they needed for solving the very first task), the mid-achieving students were split up and categorized in either of the two other groups: higher- and lower-achieving students. The final distribution of the students in different conditions is presented in the Results section below.

All students played the game at their own pace. Some finished all tasks before the end of the third lesson and others did not complete all tasks. At the end of the last lesson the students performed a post-test, consisting of six multiple-choice questions (see Silvervarg et al., 2021). After that, the students were asked to complete the 'agent opinion' questionnaire (previously used in Silvervarg and Månsson (2018)), assessing their opinion of Timy by answering four statements: The time-elf… (i) is like a classmate, (ii) is like an assistant, (iii) is like a younger sibling, and (iv) needs my help. Each statement could be answered on a continuous bipolar scale ranging from 'completely disagree' to 'completely agree'. The last question, whether Timy needs help or not, relates to the protégé effect, i.e., that you put more effort into a task when teaching someone else compared to when learning for oneself (Chase et al., 2009).

The 'agent opinion' questionnaire was distributed to participants in both conditions ('Timy as TA' and 'Timy as narrator'), wherefore the results also serve as a validation of the intervention by providing information on the extent to which the time-elf was perceived as intended.

3. Results

The results are based on the dataset of 156 students described above. All of these students completed at least one historical mission in the game, responded to the post-test, and filled in the 'agent opinion' questionnaire. For the analyses, the students were distributed into four groups based on their achievement level (higher/lower, described in the method section above) and the experimental condition they were assigned to (Timy as Narrator; Timy as TA), see Table 1.

Table 1. *Distribution of students within different conditions*

Achievement Level	Experimental Condition	*N*
Lower	Timy as Narrator	37
Lower	Timy as TA	39
Higher	Timy as Narrator	37
Higher	Timy as TA	43
	Σ	156

In the following sections we first investigate the students' characterization of the time-elf (Timy), and how they evaluated Timy's need for help. After that comes an evaluation of the condition with 'Timy as a TA' (in comparison with 'Timy as a narrator'), and finally we investigate the relationship between the individual student's opinion about Timy's need for help and his or her performance.

3.1 RQ.1: The students' characterization of the time-elf Timy and their thoughts about Timy's need for help

The students' opinion of Timy, given by their ratings of the four statements (The time-elf… (i) is like a classmate, (ii) is like an assistant, (iii) is like a younger sibling, (iv) needs my help), was evaluated by measuring the participants' responses in millimeters along a 100 mm bipolar analogous scale, and then transforming this measure to a value between 0 and 1. The transformed results of the responses, presented in Table 2 below, reveal that the students' opinion of Timy differed significantly between the two conditions for all statements except 'classmate'. This is especially notable for the students' ratings on the statement "The time-elf needs my help" ('Timy as TA': $M = 0.68$, $SD = 0.31$, 'Timy as Narrator': $M = 0.38$, $SD = 0.33$), (Welch t-test: $t(150.5) = -5.88$, $p << .001$, $d = 0.95$)), where the means from the two conditions ended up on different sides of the 0.50 mid-value on the rating scale. This suggests that Timy's need for help worked on a conscious level and supports the notion that the students in the TA-condition were aware of the TA's role as a tutee and that they were supposed to help him/her.

The students' perception of the time-elf as an 'assistant' also differed significantly between conditions although both means were below the 0.5 mid-point, indicating that none of the groups considered this statement to be an appropriate description of Timy. When it comes to the other two categories: 'classmate' and 'younger sibling', there was no or only a small significant effect between conditions. Both statements also fell below the 0.5 mid-point; especially so the 'younger sibling' statement. Clearly, the time-elf Timy did not fit well with neither of these three statements.

Table 2. *The students' opinion, Mean (M; Range: 0–1) and Standard Deviation (SD), about the time-elf Timy for the two conditions × four statements (above) and Welch two sample t-tests between conditions for the four statements (below).*

Condition	***M* (*SD*) for Timy as…**			
	classmate	**assistant**	**younger sibling**	**in need of my help**
Timy as narrator	0.37 (0.27)	0.43 (0.28)	0.16 (0.24)	0.38 (0.33)
Timy as TA	0.32 (0.32)	0.26 (0.29)	0.28 (0.32)	0.68 (0.31)
Welch two sample *t*-test				
df	150.7	152.0	147.4	150.5
t-value	1.05	3.68	−2.55	−5.88
p-value[1]	0.30	< 0.001 ***	0.024 *	<<0.001 ***

[1] *p*-values Holm corrected for family-wise error rate.

3.2 The impact of the teachable agent on performance

To ensure that the TA had a significant effect on learning, we started out with a comparison of performance levels for different student groups and conditions (even though this is not articulated as a research question). This was done by classifying the students as 'good performing' or 'not-as-good performing' (depending on their results on tasks in the game and scores on the post-test) and setting up a logistic regression model to predict the probabilities for each one of the students to fall into the 'good performing' category. The following variables were used:

- **Good performing**: *GoodPerf* (binary dependent variable). Defined as having at least 50% correct on the post-test and completing at least three tasks; classified as 1 if the student passed this threshold, otherwise 0.
- **Student achievement level**: *AchLev* (categorical independent variable). Based on teacher assessments on reading proficiency an initial in-game performance. Two levels: higher and lower [Higher, Lower]
- **Time-elf condition**: *CndElf* (categorical independent variable). Two conditions: Timy as a TA and Timy as Narrator [TA, Narrator].
- **Interaction effects**: It was hypothesized that there might be an interaction effect between the two independent variables.

The following binomial logistic regression model predicting *GoodPerf* was then set up:

$$\text{logit}(GoodPerf) = \beta_0 + \beta_1 AchLev + \beta_2 CndElf + \beta_3 CndElf{:}AchLev$$

The best model fit was found by backward and forward stepwise elimination, revealing that all predictors had a significant contribution. This model performed significantly better than an intercept-only base line model ($\chi^2(3)$: 34.0, $p < .001$), having a reasonable fit (C-statistics: 0.77, Somers' D_{XY}: 0.54). The logistic regression analysis reveals that both the student achievement level (*AchLev*) and the interaction effect between *AchLev* and *CndElf* significantly predicted students as being 'good performing' (*GoodPerf*). See Table 3 for more details and statistics. The predicted probabilities for students within different conditions and achievement levels being 'good performing' are presented in Figure 3 below.

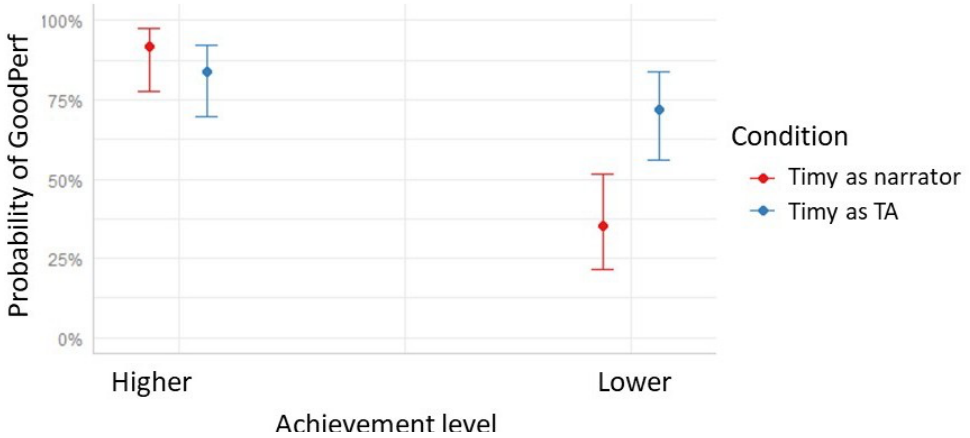

Figure 3. Probability of students being 'good-performing' (based on in-game performance and post-test scores) in relation to condition (*Timy as a TA* and *Timy as Narrator)* and achievement level (*Higher/Lower*).

Table 3. *Summary of the binomial logistic regression model fitted to predict students as 'Good Performing' (N = 156)*

Predictors	Coeff.	Odds.Ratio	Std. Err.	z-value	Pr (>|z|)	Sign.
Intercept	2.43	11.36	0.60	4.03	< 0.001	***
AchLev[Lower]	−3.04	0.05	0.69	−4.83	< 0.001	***
CndElf[TA]	−0.79	0.45	0.73	1.08	0.28	
AchLev[Lower]:*CndElf*[TA]	2.34	10.38	0.88	2.64	< 0.01	**

Model statistics. AIC: 161, C-value: 0.77, Somers' D_{XY}: 0.54, Likelihood ratio test: $\chi^2(3)$: 34.04, $p < .001$

Significance codes: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.3 RQ.2: Opinions about Timy as a TA for students with different achievement levels

Since the TA, in line with previous studies, had a significant impact on lower-achieving – but not higher-achieving – students (Silvervarg et al., 2021), we wanted to investigate if this could be related to differences in the students' opinions about their TA. A first evaluation was done by, for each student in the 'Timy as TA' group ($N = 82$), assigning Timy to one of the four statements in the 'agent opinion' questionnaire. That is, the statement that received the largest value was assumed to correspond to the individual student's 'best' description of Timy, as long as it was above the mid-point value of 0.5. Otherwise, Timy was classified as a 'NoType'. The results clearly suggest that within this condition, higher- and lower-achieving students have similar opinions about Timy (Chi-squared test: $\chi^2(9, N = 82) = 4.27$, $p = .37$).

Table 4. *Characterization of Timy within the group 'Timy as a TA' (N = 82) for higher- and lower achieving students*

Achievement Level	Tutee	Classmate	Assistant	Younger Sibling	No Type	*n*
Higher	27	5	1	2	8	43
Lower	19	5	5	1	9	39

Thereafter, we compared the two student groups with respect to the 'agent opinion' questionnaire measure for 'The time-elf [Timy] needs my help' (Higher: $M = 0.74$, $SD = 0.25$, Lower: $M = 0.63$, $SD = 0.36$). Here again there weren't any significant difference between the ratings from students with different achievement levels (Welch t-test: $t(67.0) = 1.54$, $p = 0.13$).

3.4 RQ.3: The impact of the students' opinion of the teachable agent's need of help on performance

To investigate Research Question 3: "*Will the student's opinion of the teachable agent's need of help influence their performance, and will such a potential impact differ between higher- and lower-achieving students?*", the students' ratings of 'Timy's need for help' (here referred to as *ElfHelp*) in the 'Timy as TA'-condition was used as a possible predictor in the following logistic regression model:

$$\text{logit}(GoodPerf) = \beta_0 + \beta_1 AchLev + \beta_2 ElfHelp + \beta_3 AchLev{:}ElfHelp$$

As described in the method section above, *ElfHelp* is a continuous independent variable, ranging between 0 and 1. The other two variables (*GoodPerf*[0, 1] and *AchLevf*[Higher, Lower]) have been described in section 3.2 above. The best model was fitted to the dataset by a backward and forward stepwise elimination and is presented in Table 5. Results show that only the variable *ElfHelp* contributed to the model and significantly predicted *GoodPerf* ($p < 0.01$) for the students (in the condition where Timy played the role as a TA). Neither achievement level (*AchLev*) nor the interaction effect had any significant contribution. The model is visualized in Figure 4 below:

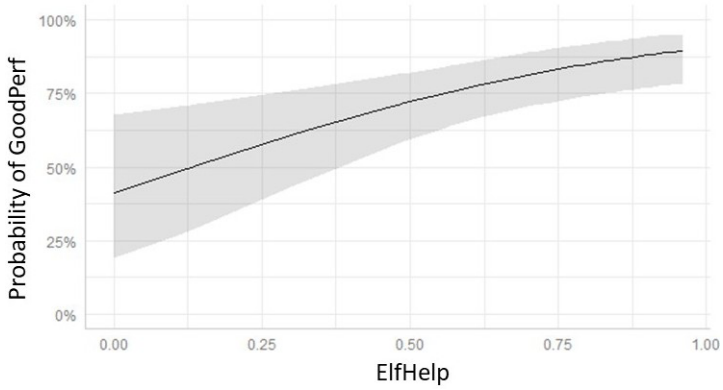

Figure 4. Probability of students in the 'Timy as TA'-condition being 'good-performing' (based on in-game performance and post-test scores) in relation to the measure *ElfHelp*.

Table 5. *Summary of the binomial logistic regression model GoodPerf ~ $\beta_0 + \beta_1 ElfHelp$ for students in the 'Timy as TA'-condition (N = 82).*

Predictors	**Coeff.**	**Odds.Ratio**	**Std. Err.**	**z-value**	**Pr (>|z|)**	**Sign.**
Intercept	−0.35	0.17	0.56	-0.62	0.54	ns
ElfHelp	2.61	13.60	0.85	3.07	< 0.01	**

Model statistics. AIC: 80, C-value: 0.72, Somers' D_{XY}: 0.44, Likelihood ratio test: $\chi^2(1)$: 10.17, $p < .01$

Significance codes: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4. Discussion

Even though a series of studies have investigated the effects of different TA personalities, students' opinions on the extent to which their TA needs their help is rarely examined or questioned. TAs are generally presumed to signal to the students to teach them, and since teaching someone else may have an effect on learning, this effect is assumed to relate to students' awareness of the TA's need of being instructed. In this study, we were interested in scrutinizing this relation and see if: (i) students' conception of an agent differ depending on whether the agent acts as a TA or not, (ii) students with different achievement levels think differently about their TAs' need for help, and (iii) the students' evaluation of the TA's need for help may have an impact on performance and learning – and if this relationship differed between higher- and lower-achieving students. In this pursuit, we ended up with

some interesting results, whereof some are in line with previous research and others are more unexpected.

First, characterizing a teachable agent – or perhaps any agent – is not straightforward. In our case, the students in the 'Timy as TA'-condition did not think of Timy as a classmate or a younger sibling, even though both of these characters can represent someone in need of help. This can be contrasted to the results in Silvervarg and Månsson (2018) where students in the experimental condition (who received additional verbal instructions about Timy's role as a tutee) were more inclined than students in the control group to characterize their TA as a classmate or a younger sibling. It might be, that by emphasizing Timy's needs of help in the beginning, the students came to care more about him/her, building a personal bond and thus also being more inclined to categorize him/her as a close relative or a friend. Interestingly, the students in the 'Timy as narrator'-condition were more inclined to rank Timy as an assistant (although the average value was below 0.5), even though s/he only introduced the game and did not provide any further help.

Our results clearly show that Timy's role in the game affected the students' opinions and thoughts about him, and the students in the condition where Timy acted as a TA also stated that he/she needed help. This was not the case if Timy only took the role as a narrator in the game. Noticeably, the students in the 'Timy as a TA'-condition were aware of their responsibilities and roles as teachers.

Not surprisingly, the condition 'Timy as TA' had a clear positive effect on performance and learning for students with lower achievement levels (assessed by combining teacher assessments on reading proficiency with initial in-game performance). This finding is in line with previous research (Chase et al., 2009; Sjöden & Gulz, 2015; Pareto et al., 2009; Tärning et al., 2017; Tärning et al., 2019). Figure 3 above also shows a minor tendency for higher-achieving students to perform less well when Timy was acting as their TA, but this effect is not significant. However, having a TA evens out the differences between lower- and higher-achieving students' learning gains.

More interestingly, the characterization of Timy and the self-report measure for Timy's need for help (within the 'Timy as TA'-condition) didn't differ significantly between students with different achievement levels. Even if our results reveal a tendency for lower-achieving students to report a slightly lower 'Timy is in need of help'-value than higher-achieving students, the difference between the groups is not significant. The reason for the different effects of the TA on learning for different student groups can thus not originate entirely from differences in opinions and thoughts about Timy as a TA.

Moreover, looking at ratings on the TAs need for help and relating them to the individual student's learning outcomes (measured by a combination of in-game performance and post-test-scores), it becomes clear that the only significant predictor of being a 'Good Performing' student is the student's thoughts about Timy's need for help (*ElfHelp*). Neither the achievement level (*AchLev*) nor the interaction effect between *AchLev* and *ElfHelp* was significant. This finding is new (and to our knowledge, not previously investigated). Since higher-achieving students often are less affected by TAs in general (Chase et al., 2009; Sjödén & Gulz, 2015; Pareto et al., 2009; Tärning et al., 2017; Tärning et al, 2019) we could assume that they also would be less affected by their personal opinions of them (in this case, how much help the time-elf Timy needs). However, this was not the case. *All* students, independently of achievement level, benefitted from perceiving the TA as someone in need of help. This could also be formulated as '*for a TA to have an effect on learning, it has to be designed in a way that the student takes its needs seriously'*. As a design consequence, the need for help ought to be clearly communicated and emphasized. In addition, taking into account the slight possibility that higher-achieving students may actually perform *worse* when working with a TA, these students also need to be convinced about the TA's needs.

Finally, the study has limitations. First of all, the ratings from the 'agent opinion' questionnaire were all gathered at the end of the experiment, after the gameplay. This means that the students might have felt differently about Timy when actually playing, and that these feelings might have varied during the game, perhaps due to various obstacles and difficulties. It may also be the case, that the self-report measure of Timy's need for help is an after-construction, and a sort of validation of the individual student's success. That is, students with stronger progression and result on the post-test, might have attributed the success to their responsibilities towards Timy and his need for help. Students reporting a lower value for Timy's need for help might, on the other hand, have 'blamed' Timy for their lack of success, avoiding taking full responsibility for their teaching (and learning).

The results of this study point to potential learning gains from designing teachable agents so that their need for help comes forth clearly. How this should be done is something to be investigated further. Should the agent, for example, specifically be asking for help and/or should more subtle behavioral patterns be exploited to communicate this? Further, the results imply that someone in need of help can trigger learning, but why is this so? Do students see the TA as a safety net in that the TA performs and takes tests, or would they prefer to play on their own, including taking the tests – and are such preferences similar or different for different groups of students?

Acknowledgements

The presented research has been funded by the Marianne and Marcus Wallenberg Foundation.

References

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, *19*(3- 4), 363-392.

Ceha, J., Lee, K. J., Nilsen, E., Goh, J., & Law, E. (2021. Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-14). New York, NY: ACM.

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of science education and technology*, *18*(4), 334-352.

Kim, Y., & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160-169.

Okita, S. Y., & Schwartz, D. L. (2013). Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *Journal of the Learning Sciences*, *22*(3), 375-412.

Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., & Cassell, J. (2012), "Oh dear stacy!" social interaction, elaboration, and learning with teachable agents. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 39-48). New York, NY: ACM.

Pareto, L., Schwartz, D. L., & Svensson, L. (2009). Learning by guiding a teachable agent to play an educational game. In V. Demitrova, R. Mizoguchi, B. du Boulay, & A. C. Graeseer (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 662-664). IOS press.

Silvervarg, A., Kirkegaard, C., Nirme, J., Haake, M., & Gulz, A. (2014). Steps towards a challenging teachable agent. In T. Bickmore, S. Marsella, & C. Sidner (Eds.), *Lecture notes in computer science: Vol. 8637. Intelligent Virtual Agents* (pp. 410-419). Springer, Cham.

Silvervarg, A., & Månsson, K. (2018). How do you introduce an agent? The effect of introduction type on how a teachable agent is experienced by students. In A. Bogdanovych, D. Richards, & S. Simoff (Eds.), *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (pp. 29-34). New York, NY: ACM.

Silvervarg, A., Wolf, R., Blair, K. P., Haake, M., & Gulz, A. (2021). How teachable agents influence students' responses to critical constructive feedback. *Journal of Research on Technology in Education*, *53*(1), 67-88.

Sjödén, B., & Gulz, A. (2015). From learning companions to testing companions: Experience with a teachable agent motivates students' performance on summative tests. In C. Conati, N. Heffernan, A. Mitrovic, & M.F. Verdejo (Eds.), *Lecture notes in computer science: Vol. 9112. Artificial Intelligence in Education* (pp. 459-469). Berlin/Heidelberg, Germany: Springer-Verlag.

Tärning, B., Gulz, A., & Haake, M. (2017). Supporting low-performing students by manipulating self-efficacy in digital tutees. In G. Gunzelmann, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 1169-1174). Austin, Texas: Cognitive Science Society.

Tärning, B., Silvervarg, A., Gulz, A., & Haake, M. (2019). Instructing a teachable agent with low or high self-efficacy – does similarity attract? *International Journal of Artificial Intelligence in Education*, *29*(1), 89-121.

Tärning, B., & Silvervarg, A. (2019). "I Didn't Understand, I′ m Really Not Very Smart" – How Design of a Digital Tutee's Self-Efficacy Affects Conversation and Student Behavior in a Digital Math Game. *Education Sciences*, *9*(3), 197-218.