

Article

# “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

Betty Tärning<sup>1</sup> and Annika Silvervarg<sup>2,\*</sup> <sup>1</sup> Lund University Cognitive Science, Lund University, 22100 Lund, Sweden<sup>2</sup> Department of Computer and Information Science, Linköping University, 58183 Linköping, Sweden

\* Correspondence: annika.silvervarg@liu.se

Received: 14 May 2019; Accepted: 5 July 2019; Published: 24 July 2019



**Abstract:** How should a pedagogical agent in educational software be designed to support student learning? This question is complex seeing as there are many types of pedagogical agents and design features, and the effect on different student groups can vary. In this paper we explore the effects of designing a pedagogical agent’s self-efficacy in order to see what effects this has on students’ interaction with it. We have analyzed chat logs from an educational math game incorporating an agent, which acts as a digital tutee. The tutee expresses high or low self-efficacy through feedback given in the chat. This has been performed in relation to the students own self-efficacy. Our previous results indicated that it is more beneficial to design a digital tutee with low self-efficacy than one with high self-efficacy. In this paper, these results are further explored and explained in terms of an increase in the protégé effect and a reverse role modelling effect, whereby the students encourage digital tutees with low self-efficacy. However, there are indications of potential drawbacks that should be further investigated. Some students expressed frustration with the digital tutee with low self-efficacy. A future direction could be to look at more adaptive agents that change their self-efficacy over time as they learn.

**Keywords:** educational math game; teachable agent; digital tutee; self-efficacy; conversational pedagogical agent; chatbot

## 1. Introduction

Educational software is now increasingly common in many schools. Some of these are built around or include pedagogical agents. A pedagogical agent is a computer character in a pedagogical role, often based on some type of artificial intelligence. Some of these pedagogical agents can be really helpful and add much to the learning experience, but some are not as beneficial. So, how should a pedagogical agent in educational software be designed to support student learning? This question is complex as there are many types of pedagogical agent which can take different roles in the learning process. There are, for example, tutor agents from which the student can learn [1], agents that work together with the student as companions or peers [2–4], agents that take an expert role [5,6], and agents that act as mentors [7,8]. Yet another pedagogical role is that of a tutee, where the agent acts as the one being taught while the real student takes the teacher role. These agents are called teachable agents [9] or digital tutees. A digital tutee is based on the idea that by teaching someone else you learn for yourself. Indeed, learning by teaching has been shown to be an efficient way to learn [10–12].

The results of many studies involving pedagogical agents indicate that they can have positive effects on learning [13–16] and on self-efficacy [17–20]. For example, Kim, Baylor and Shen [21] showed

that girls who interacted with a pedagogical agent in an educational math game developed a more positive attitude towards mathematics and increased their self-efficacy beliefs in the subject compared to girls who played the same game but without an agent. Similarly [20] found that girls increased their self-efficacy beliefs in learning mathematics after working with an animated agent embedded in computer-based learning. In [19] they found that third graders who taught a digital tutee for nine weeks showed a significantly larger gain in self-efficacy compared to the control group who had engaged in regular math classes.

However, the mere addition of a pedagogical agent to a learning environment does not automatically improve learning or provide other beneficial effects. There are several aspects that need to be carefully considered when designing a pedagogical agent: visual appearance [22–24], how the pedagogical agent behaves and interacts [8,21,25,26], and other characteristics such as whether the agent has high or low domain competence [3,4,8,17,27,28]. For example, [4] explored the agent's competence (high vs. low) in combination with different interaction styles (proactive vs. responsive). They found that students who interacted with a peer agent with high competence were better at applying what they had learned and showed a more positive attitude towards their agent. On the other hand, students interacting with a peer agent with low competence showed an increase in self-efficacy. An increase in self-efficacy was also found for students who worked with a more responsive agent. Hietala and Niemirepo [3] similarly found that students in general preferred to collaborate with a more competent digital peer compared to a weak digital peer when presented with a choice. Uresti [27] Presents opposite results pointing towards a trend (although not significant) that students who interacted with a weak learning companion learned more than students who interacted with a more competent companion. Therefore, even though the evidence is not conclusive as to whether an agent with high or low competence is most beneficial, we know that pedagogical agents with a high or low competence can have an effect on students' self-efficacy and learning. In this paper, we explore these themes further, focusing on how the self-efficacy of both agent and student affect how students interact with the agent.

### 1.1. Conversational Teachable Agents

The type of pedagogical agent used in the study presented in this paper is a teachable conversational agent. It is teachable in the sense that it knows nothing about the topic from the beginning but then learns from the student, who acts as the teacher. By teaching the agent, the student at the same time learns for herself. This effect has been proven efficient in human–human interaction. Moreover, the same effect is also present in human–agent interaction [29,30]. Chase and colleagues [30] found that students who were asked to learn in order to teach a digital tutee put more effort into the learning task compared to when students were asked to learn in order to take a test for themselves. This difference in effort and engagement is referred to as the protégé effect. Not only did the students in the study put more effort into the task when they were later supposed to teach, they also learned more in the end. Having a protégé (such as a digital tutee) can thus increase the motivation for learning. In addition, Chase et al. [30] proposes that teaching a digital tutee can offer what they term an ego protective buffer: that is, the digital tutee protects them from the experience of direct failure, since it is the tutee that fails at a task or a test—even though students are generally aware of the fact that the (un)success of the tutee reflects their own teaching of it. Nevertheless, the failure can be shared with the tutee, which then shields the student from forming negative thoughts about themselves and their accomplishments.

Sjödén and colleagues [31] made another observation regarding the social relation a student can build with her digital tutee. They found that low-performing students improved dramatically on a posttest when they had their digital tutee present during testing compared to when they did not. This difference was not found for high performing students. Even though the digital tutee did not contribute with anything but its mere presence, this had a positive effect for low performing students.

With respect to the learning by teaching paradigm, the kind of feedback students receive when using such software needs to be highlighted. In most studies on feedback and learning, feedback is something that is provided to the student from the teacher and concerns the students' performance. In

the teachable agent paradigm, the direction of the feedback is different. Here, it is the teacher (i.e., the real student) that receives feedback on how well they have been teaching by observing how well their tutee (i.e., the digital tutee) performs. The digital tutee provides feedback to the student regarding its (the tutee's) own ability to solve tasks without explicitly saying anything about the student's teaching abilities. Implicitly, however, a student can make use of the feedback from the digital tutee, including information on how the tutee performs, to infer how much she herself knows/or how well she has taught her tutee. This type of feedback is what [29] call recursive feedback, namely feedback that occurs when the tutor observes her students use what she has taught them. This type of recursive feedback is present in the math game used in this study. It appears when the digital tutee attempts to play independently using its knowledge regarding the rules and strategies of the game as learned from the student. However, it also appears in the chat dialogue when the digital tutee reflects upon its own learning and performance. These reflections make up our manipulation in that they are colored by the self-efficacy the digital tutee is assigned (high or low).

Our digital tutee also belongs to the pedagogical agent subgroup called conversational pedagogical agents. Conversational agents in the area of education are primarily able to carry out conversations relating to the learning topic at hand. However, some of them are also able to carry out so-called off-task conversation or "small-talk", not related to the learning topic as such. Off-task conversation can make a learning situation more relaxed and has been shown to promote trust and rapport-building [32,33]. Off-task conversation is also something that many students are familiar with from real world learning experiences. Classroom interactions encompass a mix of on-task and off-task interactions and the teacher does not just go on about the topic to be learned, usually there is an ongoing conversation with little (apparent) relation to the topic to be learned. To be noted is that not all students experience off-task conversation as something positive; some find it time-consuming and meaningless [34].

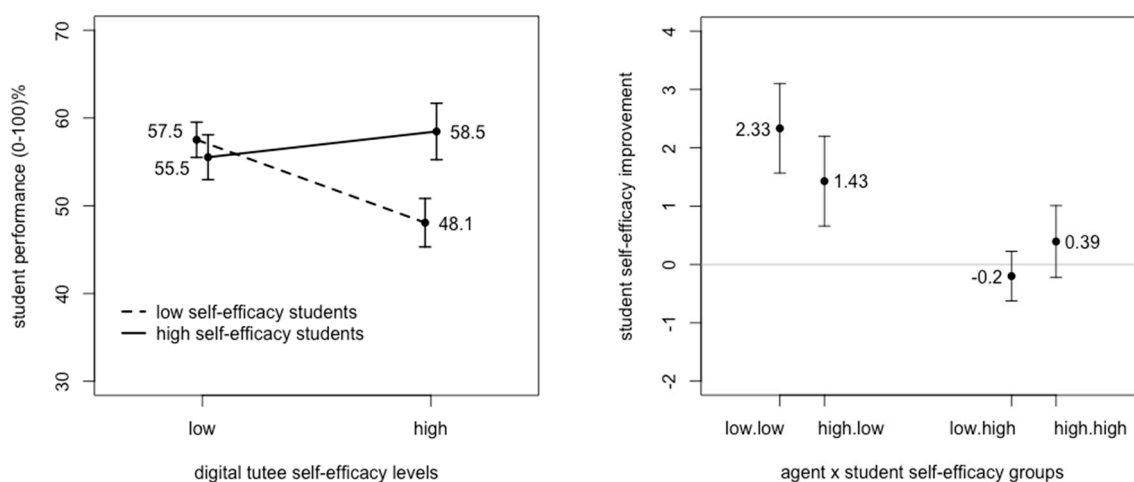
Previous research with the educational game used in this study investigated the effects of the off-task conversation within the chat. Those results showed that overall, the students did not experience the off-task conversation as disturbing, and students who were allowed to engage in off-task conversation had a more positive game experience compared to students who did not have the opportunity [35]. The study also explored whether high-, mid-, and low-achievers would differ with respect to their experience of the off-task conversational module (the chat module). The outcome was that high- and mid-achievers liked the software more when the off-task conversation (the chat) was included—but that they chose to chat less than the low-achievers. Conversely, low-achievers were more indifferent towards the chat—but they chatted more than high- and mid-achievers. In a follow-up analysis of the material [36] found that, the engagement differed between these sets of students. High-achieving students showed greater engagement than the low-achieving students did when chatting, but in situations where they appeared unengaged in the chat, they tended to choose to quit the chat and refrain from starting a new. The low achievers on the other hand were more inclined to continue a chat even when they appeared disengaged. The authors speculate that low-achievers do not take control over their learning situation to the same extent as high-achievers.

### *1.2. Designing a Teachable Agent with High or Low Self-Efficacy*

From previous studies we know that pedagogical agents can have beneficial effects on for example students learning experience and self-efficacy. By for example manipulating the agent's competence some students are more or less willing to work with it [3,4,27]. In the context of a teachable agent, the level of competence or expertise is not an easily manipulated variable, since the competence of the digital tutee reflects the real student's teaching. Simply put, if the student teaches the digital tutee well, the tutee will learn and increase its knowledge and competence. On the other hand, if the student does not teach her digital tutee well, the tutee will not increase its knowledge and competence. In contrast, the characteristic of self-efficacy is possible to design and manipulate in a digital tutee and this is what we have done.

In [37], we studied whether the manipulation of self-efficacy in a digital tutee—in terms of low versus high self-efficacy—would affect any of the following for the (real) students who acted as teachers for the digital tutee: (i) their self-efficacy (ii) their in-game performance, (iii) their attitude towards the digital tutee. The study made use of an educational game targeting mathematics and the base ten concept [38], further described in Section 2.1. The digital tutee interacted with the student both via a scripted multiple-choice conversation and via a natural language chat conversation, see Section 2.1.1. In the chat conversation in which the digital tutee commented on her performance, expectations and ability to perform and learn, the tutees' self-efficacy was manipulated to be low or high. Following previous research focusing on matching/mis-matching effects of characteristics in students and pedagogical agents (e.g., [3]), we investigated possible effects of the digital tutee and student having similar or dissimilar self-efficacy. A matching pair was when the digital tutee and the student both had low or high self-efficacy regarding mathematics and a mis-matching pair when the agent had high self-efficacy while the student had low self-efficacy and vice versa.

The analysis in [37] showed that interacting with a digital tutee with low self-efficacy was beneficial for students' performance. This was especially apparent for students who themselves had had reported low self-efficacy, they significantly increased their performance and performed as well as students with high self-efficacy when interacting with a digital tutee with low self-efficacy, see Figure 1 (left side). Furthermore, the digital tutee with low self-efficacy also had another positive effect on students with low self-efficacy, who significantly increased their self-efficacy when interacting with it, see Figure 1 (right side).



**Figure 1.** On the Left: Student in-game performance with regard to digital tutee vs. student self-efficacy. On the Right: Student improvement in self-efficacy with regard to digital tutee vs. student self-efficacy.

Together, these results indicate that interacting with a teachable agent with low self-efficacy was beneficial for students overall and in particular for students with low self-efficacy. In [37] we speculated that this can be due to the protégé effect, in that students make a greater effort for an agent lacking in self-efficacy that seems more in need of help than an agent with high self-efficacy. The ego-protective buffer can also be in play, in that students with low self-efficacy, that also tend to be low performing, do not perceive the feedback as negatively since it is recursive and aimed at the agent rather than them.

In order to make these results into guidelines on how to design pedagogical agents with low (or possibly high) self-efficacy and know the benefits and possible disadvantages of doing so, we need to know more about the possible causes of these effects, but also look more carefully after differences between and within different student groups. It is possible that some students do not respond well to the agent with low self-efficacy, or that the positive effect is due to different causes for different groups. To do this, we turn to the chat dialogues between students and their digital tutees, that were collected during the study reported in [37]. We wanted to explore if we could find any patterns in the

chat interaction between tutee and student that possibly further explained the results in [37]. In this paper we explore potential differences in how the students responded to the feedback that expressed either high or low self-efficacy of the agent, but also if the students on their own initiative commented on the agent's intelligence and competence, or its attitude. We also compare matched and mis-matched cases, i.e., where students had similar or dissimilar self-efficacy (low or high) as their digital tutee. More specifically the research questions are the following:

Q1. What difference does the self-efficacy of digital tutees and students make, to what extent, and how:

- (a) the students react and respond to the digital tutees' feedback?
- (b) the students comment on the digital tutees' intelligence and competence?
- (c) the students comment on the digital tutees' attitude?

Q2. Are there any relations between students' chat behavior and students' performance?

## 2. Materials and Methods

### 2.1. The Math Game

The math game "The Squares Family" [38] trains basic arithmetic skills related to the base-ten concept, by means of different board games, see Figure 2. Instead of using numbers, the game uses blocks and boxes to visualize the base-ten concept. These block and boxes can be placed on the game board partitions. On the dark blue board, a maximum of nine red boxes can be placed within the ropes, on the light blue area nine orange boxes, and on the green area yellow boxes. If there are nine red boxes in the dark blue area and a player chooses a card which adds one more, these ten red boxes will be packed into one orange box and moved to the light blue area. Similarly ten orange boxes will be packed into one yellow and moved to the green area if there are more than ten after a card is played. Thus, red squares represent ones, orange boxes tens, and yellow boxes represents hundreds. The board is in the middle and two sets of cards, one for each player (in this case a student vs. the computer). The players take turn choosing one of their cards, the content of which is added to (or subtracted from) the board. In Figure 2 the student Annika has chosen the card 50, i.e., five orange boxes. A star is awarded for each carry-over, e.g., when ten red squares are transformed into an orange box, and the player with the most stars at the end wins.

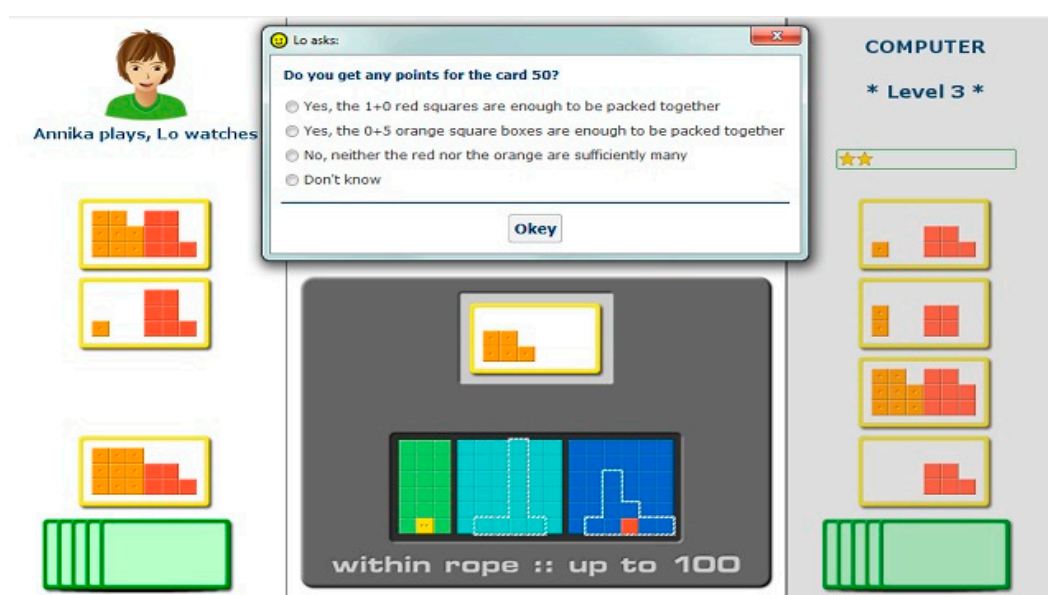


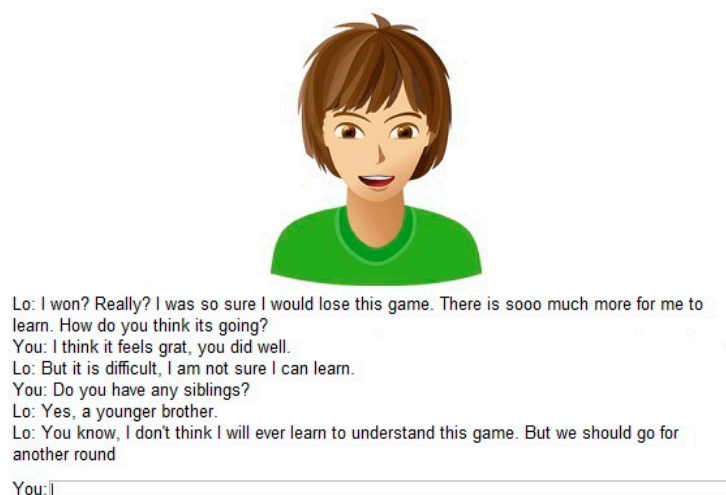
Figure 2. The game in Observe mode, Lo is asking a multiple choice question.

The game also incorporates a digital tutee, named Lo (see Figure 2), whom the student should teach the base ten concept. The agent learns the rules of the game and the underlying mathematical model through observing the student's actions and by asking multiple choice questions. The student can teach Lo in two different modes. In the "observe" mode the student plays and the digital tutee learns by watching and by asking multiple choice questions to the student (see Figure 1). In the "try" mode the digital tutee suggests which card to choose but can be corrected by the student who has the possibility to pick another possibly better card. In both of these modes the digital tutee regularly asks multiple choice questions regarding the game and the underlying math model. The questions always relate to the current game situation and most often the choice(s) of card(s) just made. There is one correct answer, two faulty answers and one "Don't know" answer.

There is also a third mode, "play" mode, where the digital tutee plays on its own based on what the student has taught it. This gives the student the opportunity to watch how well its digital tutee performs. The agent's performance is a direct reflection of how well the student have understood the game and the underlying mathematical model.

### 2.1.1. The Chat

The chat is where the manipulation of the digital tutee's self-efficacy takes place. The chat appeared after every finished game, besides from when the student plays alone without their digital tutee, which only happened in the beginning when the student was getting to know the game. The chat always started with a feedback- sentence that expressed the digital tutees self-efficacy (high or low) for example "I have learnt a lot really quickly. I think I will have learned everything very soon" (high self-efficacy) or "It felt like I did not understand everything we went through during this round, I'm really not that smart" (low self-efficacy). This was followed by a question to the student, "*How do you think it's going?*". After that a free format chat followed for one minute, that was automatically terminated with a statement from the agent that also reflected its self-efficacy, for example "You know, I don't think I will ever learn to understand this game. But should we go for another round" (see Figure 3 for an example of a typical chat dialogue).



**Figure 3.** An example from chatting with Lo when displaying low self-efficacy.

Within the chat the students were free to talk about whatever they wanted to, for example, on-task topics such as school, math, the game or learning in general as well as off-task topics such hobbies, music and movies. In the current version, the digital tutee is able to handle greetings, ask and respond to questions and statements on various topics, ask and respond to follow up questions, and to tell mini-narratives, as illustrated in Figure 4.

```

User: Hi
Agent: Hi
User: Hello
Agent: What type of music do you like?
User: much u?
Agent: everything but dance band music hehe
User: hehe
Agent: lol
User: lol
Agent: I listen to rock and rnb I think that's the best music

```

**Figure 4.** An example chat with greetings, questions and follow-up questions.

The chat allowed for mixed-initiative dialogue, which means that both the digital tutee and the student could take the initiative and ask questions. The student could ignore a question from the tutee and instead pose a new question. When the digital tutee did not understand, it had a strategy where it first asked for a clarification, then made a general request to change topic, and thirdly suggested a novel topic, as illustrated in Figure 5.

```

User: fgd
Agent: I did not understand that
User: cb
Agent: I did not catch that, can we talk about something else?
User: s
Agent: Let's talk about school
User: s
Agent: My school is okay, a little too much homework, wbu?

```

**Figure 5.** An example chat where the tutee does not understand the student's utterances.

#### Feedback Sentences

One of the authors constructed the feedback sentences and pilot tested them on 22 fourth graders who were not a part of the study. The students in the pilot study were asked to read the sentences (presented in a randomized order) in order to evaluate whether they reflected high or low self-efficacy. They were asked to judge whether each sentence sounded like something being said by someone that was confident, not confident, or neither. The sentences that were not considered as either high or low in self-efficacy were then adjusted to match the approved sentences. Overall, there were 136 different sentences, 68 portraying Lo with high self-efficacy and 68 portraying Lo with low self-efficacy.

Since the game itself has three different modes ('Observe', 'Try' and 'Play') the sentences also needed to correspond to these three modes. In the observe mode the student plays him/herself and the digital tutee is just learning from observing. For example, a sentence that appeared after a game in observe mode could say "I'm learning the rules slowly, I'm not such a brilliant student" (digital tutee with low self-efficacy). All sentences in this mode were expressed in a first-person perspective. 'I', since the digital tutee only observed what the student did.

In the try-mode (where the digital tutee could try for herself with the student correcting if they felt the digital tutee chose the wrong card), the digital tutee could express sentences in both the I- and

we-form, for example “That’s great! I was sure that we were going to win, I think we played really well” (digital tutee with high self-efficacy).

In the last mode (play) the student was not actively participating, instead the digital tutee played herself while the student was watching. After a game in this mode the sentences were again expressed in first person for example “Boohoo, how could I lose?! I played awesomely and I chose the best cards” (digital tutee with high self-efficacy).

Each game mode was in turn divided into subcategories: ‘game result + gameplay’, ‘game result + learning’ and ‘game result + agent knowledge’. Each sentence started with a comment on the outcome of the previously played round of game—victory, defeat or even (i.e., ‘game result’). ‘Gameplay’ refers to how well Lo thought she had played “That’s awesome, we won since we choose the best cards the whole time” (high self-efficacy). ‘Learning’ reflected how much she thought she had learned during the previously round “I really didn’t learn much this round, but maybe that was not so unexpected” (low self-efficacy), and “agent knowledge” how much she thought she knew about the game in total “Wahoo, I won! But that was not so unexpected considering how good I am and how much I have learned by now” (high self-efficacy). However, in game-mode, observe there was no ‘game result + gameplay’ because the tutee was not involved in playing but only observed, and in game-mode ‘Play’ there was no ‘game result + learning’ since the tutee did not learn anything from the student in that mode.

The chat always ended with a sentence from Lo regarding her thoughts about the upcoming round, for example “I have a feeling that the next round will go really good, let’s go!” (high self-efficacy) or “I don’t feel like I understand anything but let’s play another round” (low self-efficacy). For a summary of feedback examples see Table 1.

**Table 1.** Examples of sentences by the digital tutee, reflecting either high or low self-efficacy.

Self-Efficacy		
Game Mode	Low	High
Observe	game result + learning	game result + learning
	“It felt like I did not understand everything we went through during this round, I am really not smart”	“It felt like I understood everything we went through during this round, I really am a genius”
	game result + agent knowledge	game result + agent knowledge
	“I haven’t learned so very much yet. I guess I have a lot more things to learn”	“I have learnt a lot quickly. I think I will have learned everything very soon”
Try	game result + gameplay	game result + gameplay
	“Did we win?! Wow, I thought we chose the wrong cards the whole time”	“We got pretty bad cards but we still won, we really play brilliantly”
	game result + learning	game result + learning
	“We lost . . . But I feel rather uncertain regarding the rules so maybe it wasn’t so strange that we didn’t win”	“We didn’t win but that was just bad luck. I feel very certain about the rules and how the game is played”
	game result + agent knowledge	game result + agent knowledge
Play	“I still don’t feel like I know anything about the game, I am glad we won!”	“I feel like I know everything about the game now, I don’t know how we could lose?!”
	game result + gameplay	game result + gameplay
	“I lost, maybe I am not so very good at choosing the right cards . . . ”	“It sucks that we lost! I was so sure we were going to win this round, I thought we played really well”
	game result + agent knowledge	game result + agent knowledge
Finishing sentences	“Did I win?! I was so sure I would lose, it feels like I have so much more to learn”	“Brilliant, I feel very certain about the rules now so this round felt really easy”
	“I don’t feel like I understand anything but let’s play another round”	“I have a feeling that the next round will go really good, let’s go!”



## 2.2. Procedure

The study was comprised of one pre-test session, seven game-playing sessions, and one post-test session. During the pre-session students took a math test and filled out a self-efficacy questionnaire. The same self-efficacy questionnaire and an additional questionnaire probing their experiences and their attitude towards the tutee were filled out at the post-session. During the seven (in between) game sessions the students played the game and interacted with their digital tutee

Within the game sessions, the students used the game individually, sitting in front of a stationary computer or a laptop (depending on schools). Each game session lasted approximately 30–40 min. The first time they were instructed to play the game themselves (without the digital tutee) in order to be acquainted with the game. When they had grasped the gist of the game, they were asked to train their digital tutee. Each student always played with the same digital tutee and therefore always got consistent feedback in that sense that they only communicated with a digital tutee that was portrayed as having low or high self-efficacy.

## 2.3. Participants

In total, 166 fourth graders (83 girls and 83 boys) participated in the data collection. They were recruited from four schools and nine classes in Southern Sweden in areas with relatively low socio-economic status and school performance below average. Students self-efficacy were assessed using a questionnaire based on [39] that was adapted to fit this study's purposes. The questionnaire (Appendix A) used the same question stem; "How good are you at solving these types of tasks?" translated into Swedish. All seven questions related to the base ten concepts since this was the topic in the game—for example "How good are you at solving these types of task?", "Which number should be in the blank  $670 - \underline{\quad} = 485$ ", or "You have the number 274, if you add 3826 will the result end in 00?". All items were graded in five steps from "not good at all" to "very good at".

The students were assigned to one of two conditions: a digital tutee that expressed high self-efficacy or a digital tutee that expressed low self-efficacy. The two groups were balanced with regard to the students' self-efficacy. Thus, the number of matched and mis-matched pairs of student and tutee with respect to self-efficacy was equal in the two conditions. For the purpose of the analysis, the students were divided into one of three groups (low, mid or high) according to the results on the self-efficacy questionnaire. The groups were adjusted so that all students with the same result were categorized in the same group. Students who belonged to the mid self-efficacy group were then removed from the analysis since we wanted to focus on the students at the extreme ends, those with the highest and lowest self-efficacy. A further nine students were removed from the analysis due to missing data or scarce attendance, and thus the study included data from 89 participants (47 girls and 42 boys). Thus, we ended up with 44 students belonging to the high self-efficacy group ( $M = 32.39$ ,  $SD = 4.13$ ), and 45 students belonging to the low self-efficacy group ( $M = 20.31$ ,  $SD = 1.83$ ). More details on how these were distributed for the student groups and tutee condition of high vs. low self-efficacy are provided in Table 2.

**Table 2.** Descriptive statistics for the four groups used for analysis, with tutee and student self-efficacy being high or low.

Student Self-Efficacy	Tutee Self-Efficacy	<i>N</i>	<i>M</i>	<i>SD</i>
Low	Low	23	20.35	4.16
Low	High	22	20.27	4.18
High	Low	23	32.30	1.69
High	High	21	32.48	2.02

#### 2.4. Ethics

This study was carried out in accordance with institutional guidelines, following the Swedish Research Council's guidelines for conduction of ethical research in the humanities and social sciences.

#### 2.5. Dependent Measures

Data collected through chat logs and data logs of game play formed the basis for the dependent measures presented below. Supplementary Materials consisting of an excel file that contains the coded chatlogs as well as the data for the measures are available online.

##### 2.5.1. Chat Measures

Based on the research questions the authors constructed a coding schema by adding some new categories to an already existing schema [40]. Categories that could account for frequency and valence (positive/negative/neutral) of the students' responses to the digital tutee's feedback were added, as well as categories for frequency and valence of students' comments on the tutee's intelligence, competence and attitude. Utterances could be coded with multiple categories. The coding categories and corresponding measures are explained below.

Both authors coded a partition of the chats in order to identify and discuss possible problems with the coding schema. These coded chats were then given as a complement to the coding schema to the four persons that coded the rest of the chats. Finally, one of the authors went over all coded chats and made minor adjustment to harmonize the coding.

##### Responses to Feedback

To measure the students' responses to the feedback provided by the tutee we used the code "AnswerToFeedback" (AFB). The valence of the reply—positive, negative or neutral—was also coded. Examples of such sentences are: "Good, you are a brilliant student" (positive), "So so, but we need to continue" (neutral), and, "Don't be so insecure, idiot" (negative).

"IgnoreFeedback" (IFB) was used when the students ignored the given feedback from the digital tutee and when they started talking about something completely different not relating to the feedback from the digital tutee. Examples of this could be that they started by asking their digital tutee something not related to the game, such as "What is your mother's name?" or "Do you like football?" Some students also replied with nonsense such as random letters.

These categories were used to compute the measures of frequency of responses to feedback (freqAFB), as well as frequency of positive feedback (freqPosFB). Both expressed as a value between 0 (%) and 100 (%).

##### Comments on the Tutee's Intelligence or Competence

Many students remarked on their digital tutee's intelligence or competence, which was coded with "CommentOrQuestionOnIntelligenceOrCompetence" (CI). It was noted if this was in a positive, negative or neutral way. Examples of these types of remarks include "You are very good at math", "Hey Lo, nice work, you won!" these would be regarded as positive remarks about the digital tutee. Examples of negative remarks could be "You such at math" and "Idiot". For this category no neutral responses were coded, probably because all comments concerned the tutees attitude and were emotionally loaded.

This category was used to compute two measures, the number of comments regarding the digital tutee's intelligence and/or competence made from students to their tutees (numCI), and the frequency of positive comments in relation to negative comments (freqPosCI).

##### Comments on the Tutee's Attitude

"CommentOnAttitude" (CA) was coded for whenever the students remarked on the digital tutee's attitude (towards the game and learning) and whether or not this was done in a positive, negative, or

neutral way. For example, “You have to believe in yourself” (positive), “You are not very kind when you say things like that” (negative) and “Don’t think so much” (neutral).

Due to sparse data, where most students had given none or only one positive or one negative comment, the frequencies of positive or negative comments on attitude were not calculated. Only the number of positive and negative comments from students to their agents was computed (*numPosCA*, *numNegCA*).

### 2.5.2. Performance Measures

We also measured how well the students performed while teaching the digital tutee, which indirectly measures their own learning and skills. This was measured in two ways; through the logging of their answers to the in-game multiple-choice questions about the game and the underlying model of the base ten concept, and how well they played.

#### Multiple-Choice Questions

Multiple choice questions were posed by the digital tutee three times during each game played in observe- or try mode, and the student could provide a correct, incorrect or a “Don’t know” answer (see Figure 1). Example questions are “How many orange square boxes are there in the 2 yellow square boxes on the game board?” and “How many red square boxes are needed to fill a yellow square box?” The answer reflects if the students understand that 10 red squares make up an orange box, and 10 orange boxes make up a yellow box.

A measure was calculated based on the percentage of correct answers in relation to incorrect answers using the formula  $(\text{percentage correct answers} - \text{percentage incorrect answers} + 100)/2$ . This resulted in a number between 0 and 100 where 100 means that the student answered all the questions correct, 0 means that all questions were answered incorrectly and 50 means that as many questions were answered correct as incorrect.

#### In-Game Performance

In-game performance was also measured in terms of the students’ quality of gameplay, as represented by the average ‘goodness value’ (0–100) of each card the student selected during a game. In short, the goodness value reflects how good the student’s choice of card is given the available options. Both the number of points the player can receive from the card and its strategic value in terms of preventing the opponent to receive points is taken into account. Importantly, even though goodness correlates with competitive outcome (winning correlates with high goodness), there are situations where the player cannot win (for example due to getting ‘bad’ cards) which can still reflect the player’s knowledge and ability to choose the best alternative from a ‘poor’ selection. In other words, the goodness value provides a measure of performance, which, over time, reflects the player’s learning progression in the game, independent of the number of wins and losses. For further details on the relationship between goodness values and game progression, we refer to [38].

### 2.6. Research Design and Data Analysis

This study employed a between subject  $2 \times 2$  factorial design, with tutee self-efficacy and student self-efficacy as the two factors. For research questions, Q1a and Q1b two-way ANOVAS were performed to investigate the effect of the independent variables on the dependent variables regarding responses to feedback and comments on the tutees’ intelligence and competence. Due to sparse data, for research question Q1c only the number of comments on the tutee’s attitude was calculated. Instead, a qualitative approach was taken where all comments were collected and grouped based on their content. For research question Q2, a correlation analysis was performed to explore if the students’ chat behavior related to students’ game performance.

### 3. Results

The aim of this paper is to explore how a digital tutee's self-efficacy, expressed as feedback given in a chat, affect students interaction and behavior towards the agent. Our analysis of the chat logs focused on to what extent and how students react and respond to the digital tutees' feedback, comment on the digital tutees' intelligence and competence, and comment on the digital tutees' attitude. We also wanted to see if there were any relations between students' chat behavior and students' in-game performance.

#### 3.1. Responses to Feedback

Since the self-efficacy of the teachable agent was expressed through feedback delivered in a chat, we posed our first question, Q1: "How do the students react and respond to the digital tutees feedback on what went on in the game?" The first step was to see if the students would acknowledge the feedback and questions from the tutee, such as for instance "What do you think about the next round?", or if they ignored this feedback from the digital tutee (freqAFB). Results were that, overall the students responded to 53% of the feedback.

A two-way ANOVA showed a small to medium sized significant main effect of the tutee's self-efficacy on frequency of response ( $F(1,88) = 3.99, p < 0.05, \eta_p^2 = 0.045$ ), where students responded more frequently to feedback from the digital tutee with low self-efficacy ( $M = 58.78, SD = 24.41$ ), than to feedback from the digital tutee with high self-efficacy ( $M = 47.74, SD = 27.10$ ). There was no main effect of student self-efficacy on frequency of response ( $F(1,88) = 1.39, p = 0.24$ ), nor an interaction effect of student and tutee self-efficacy ( $F(1,88) = 0.245, p = 0.62$ ), see Table 3 for means and standard deviation for these groups.

**Table 3.** Mean and standard deviation,  $M(SD)$ , for frequency of response to the digital tutees feedback.

	Tutee with High Self-Efficacy	Tutee with Low Self-Efficacy
Student with high self-efficacy	52.43 (25.95)	60.35 (19.12)
Student with low self-efficacy	43.27 (28.02)	56.91 (29.09)

The next step was to look at the cases where the student had actually responded to the feedback and the digital tutee's question, formulated as for example; "How do you feel?" When responding the student could do it in either a positive way such as "It feels very well, you did very well", or a negative way such as "It doesn't go very well, you need to practice more", or in a neutral way, writing for example "okay". Results show that on average, 72% the responses were positive (freqPosFB).

A two-way ANOVA showed a significant small to medium sized main effect ( $F(1,87) = 4.87, p < 0.05, \eta_p^2 = 0.055$ ) between students with high self-efficacy who responded positively more often ( $M = 78.55, SD = 24.46$ ), than students with low self-efficacy ( $M = 65.34, SD = 30.85$ ). There were no main effects of tutee self-efficacy ( $F(1,87) = 0.23, p = 0.63$ ) nor any interaction effect ( $F(1,87) = 0.59, p = 0.30$ ). See Table 4 for means and standard deviation for these groups.

**Table 4.** Mean and standard deviation,  $M(SD)$ , for frequency of positive responses to the digital tutees feedback.

	Tutee with High Self-Efficacy	Tutee with Low Self-Efficacy
Student with high self-efficacy	78.76 (28.43)	78.35 (20.85)
Student with low self-efficacy	62.27 (35.54)	68.41 (25.80)

The most frequent type of positive answers from the students was simply a "good" or "well" when answering the digital tutee how well they thought it proceeded. These replies accounted for approximate one third of all positive answers. Some were more superlative like "great" and "awesome", but these were rather few. One out of six answers commented on the tutee's intelligence or competence, the most frequent answers being of the type "You are good", or "You are learning", or more seldom,

“You are clever”. The neutral answers were usually a “don’t know”, “so-so” or “ok”. There were also occasions where the student instructed the digital tutee to observe more carefully or put more effort into the next round.

Frequent negative answers when the digital tutee asked, “How do you think it’s going?” were “badly” or “really badly”. Almost half of the negative answers were derogatory, or even abusive comments about the tutee’s intelligence or competence, like “You suck”, “You are dumb” or “You lost, idiot”.

To conclude question “Q1: To what extent and how do the students react and respond to the digital tutees feedback on what went on in the game?” we note that students responded more frequently to feedback from the digital tutee with low self-efficacy, and that the responses were mostly positive. Thus, the trait of having low self-efficacy in a digital tutee can lead the students to put more effort into the interaction and to respond more positively to comments about the tutees self-efficacy, learning and performance.

### 3.2. Comments on the Tutees’ Intelligence and Competence

Next, we looked at research question “Q2: To what extent and how do the students comment on the digital tutees’ intelligence and competence?” The comments in question appeared in the free conversation following the feedback from the digital tutee. Some of them were a direct response to the feedback utterance and prompted by the tutee, but more than half of these comments came spontaneously later in the conversation. Almost all students made comments, as they occurred for 89% of the students overall, for 93% of the students interacting with a tutee with low self-efficacy, and 84% of the students interacting with a tutee with high self-efficacy. On average, each student gave 4.9 comments to their digital tutee, 5.7 for students interacting with a tutee with low self-efficacy and 4.0 comments for students interacting with a tutee with high self-efficacy (numCI).

Overall the comments regarding the digital tutees intelligence or competence were negative, on average 60% of the comments were negative and only 40% were positive (freqPosCI). A two-way ANOVA showed a significant medium sized main effect of the tutee’s self-efficacy ( $F(1,78) = 5.71, p < 0.05, \eta_p^2 = 0.071$ ), where the digital tutee with low self-efficacy on average received more positive comments ( $M = 49.91, SD = 37.26$ ), than the digital tutee with high self-efficacy ( $M = 28.97, SD = 38.13$ ). There was no main effect of student self-efficacy ( $F(1,78) = 0.34, p = 0.56$ ) nor any interaction effect ( $F(1,78) = 0.21, p = 0.65$ ). See Table 5 for the mean and standard deviation of these groups.

**Table 5.** Mean and standard deviation,  $M(SD)$ , for frequency of positive comments on the digital tutees’ intelligence or competence.

	Tutee with High Self-Efficacy	Tutee with Low Self-Efficacy
Student with high self-efficacy	33.79 (36.98)	50.45 (37.30)
Student with low self-efficacy	24.67 (39.62)	49.35 (39.46)

Most of the negative comments involved saying that the digital tutee was an idiot or that (s)he sucked. However, some of the comments referred to the tutee’s abilities to learn math and the game, for example, “You are not very good at math” or things like “How can you be so stupid” and “I mean, do you actually have a brain?”. The positive comments mostly concerned the tutee’s performance and ability to play, the student saying things like “You are super good” or “You did very well”. However, students also expressed happiness regarding their digital tutees’ performance saying things like “It feels very nice when you play as good as you do” or “Oh my God, you are really good, that is so fun to see!!!”

Thus, the results for “Q1b: To what extent and how do the students comment on the digital tutees intelligence and competence?” are in line with the results on Q1a. The digital tutee with low-self efficacy received more positive comments about its intelligence and competence than the digital tutee with high self-efficacy from both students with low and high self-efficacy.

### 3.3. Comments on the Tutees' Attitude

Of special interest was to see if the students commented on the digital tutees attitude towards its learning and performance, since this attitude relates to the self-efficacy that the tutee expressed. Thus, we explored "Q1c: To what extent and how do the students comment on the digital tutees attitude?" Our results show that only 21 out of 89 students made any comments regarding the digital tutees attitude. Out of these, 17 directed the comments to the digital tutee with low self-efficacy. In other words, for the digital tutee with low self-efficacy 17 out of 45 tutees received comments, while for the digital tutee with high self-efficacy only 4 out of 44 tutees received comments on its attitude. Equally, as many students with high self-efficacy (10) as students with low self-efficacy (11) provided these comments.

Of the four comments to the tutee with high self-efficacy one was negative and expressed a frustration with the mismatch between performance and the tutees attitude "What do you mean, "you have learned a lot, I lost". Two were positive "It's good that you are confident, it will go well", "It will go well, just believe in yourself" and one was neutral "Ok, but we need to continue working".

The comments to the digital tutee with low self-efficacy are all listed in Table 6 with the exception of similar comments from the same student in the same chat session. It is noted whether a student with low or high self-efficacy gave the comment.

**Table 6.** Comments to the TA with low self-efficacy regarding its attitude from students with both low and high self-efficacy.

Student Self-Efficacy	Comment
High	You could say well
High	Why do you think so negatively
Low	Yes, but you have to say something positive too
Low	You shouldn't be so fucking negative
Low	I know, but you are pretty good just think positively
Low	Don't be unsure you idiot
Low	But you should be, idiot
Low	Do not be unsure you will win
High	Good you did that well just stop being so unsure
Low	Lo it will be fine just relax
Low	Do not worry you will do it!
Low	It's cool, Lo
High	Good if you believe in yourself, I think you can do it
High	Tell yourself you can win
High	I think you can do it!
Low	You need to focus more, do what you should, do not think of anything else!
Low	I don't like it when you say so
High	Why do you ask when you score points all the time
High	I won with 21-8, what's wrong with you
Low	Good, but you will probably say that it was not good
Low	What do you talk about, it went really well
Low	You are super GOOD DON'T YOU GET IT????
High	Yes, I think it goes well for both of us, why don't you
Low	I got 11 stars, it's not nice to say that

Out of the comments to the tutee with low self-efficacy more than half state that the student and/or tutee is doing well. However, many of these comments express some frustration or sadness over the tutees negative attitude, for example "Yes, I think it goes well for both of us, why don't you". But there are also many that are encouraging, trying to boost the tutee, for example "Don't worry you will do it". Overall many of the comments tell the tutee to be less negative and more positive (e.g., "I know, but you are rather good, just think positive"), to not be unsure (e.g., "Do not be unsure you will win") and believe in itself (e.g., "Tell yourself you can win"). Some students also tell the digital tutee to relax (e.g.,

“It’s cool, Lo”), focus on the task (e.g., “You need to focus more, do what you should, don’t think of anything else!”), or that it not kind to be so negative (e.g., “I don’t like it when you say so”). These comments vary in tone, with some being rather harsh (e.g., “You shouldn’t be so fucking negative, don’t be unsure you idiot”) and some very encouraging (e.g., “Don’t worry you will do it! Good if you believe in yourself, I think you can do it”)

The sparse data makes it hard to draw any definite conclusions, but it is important to note that while some students encourage the tutee when it expresses a low self-efficacy, some students also get a bit frustrated with it, especially if they think they or the tutee is performing well. Here, there likely are differences for students with high and low self-efficacy. Since students with high self-efficacy generally perform and teach their tutees better there will be a mismatch between the tutees’ low self-efficacy and high performance, which can lead to frustration. For students with low self-efficacy of which many will also not teach their tutees equally well, the discrepancy between the self-efficacy and performance of the tutee will not be as obvious.

### 3.4. Relations between Chat and Performance Measures

Thus far, we have found differences in how students with high and low self-efficacy interact with digital tutees expressing high or low self-efficacy in the chat. Since our starting point was to explore the effect that students with low self-efficacy perform better as well as gain a higher self-efficacy belief when interacting with a digital tutee displayed as having low rather than high self-efficacy, our final analysis concerned Q2: Are there any relations between students’ chat behavior and students’ in-game performance?

Students’ performance was calculated in two ways (see Section 2.5.2 Performance measures): (i) the proportion of correct and incorrect answers given by the student in relation to the multiple-choice questions posed by the digital tutee regarding the game and its underlying mathematical model and (ii) the goodness of cards chosen by the student during gameplay. The first measure is more directly related to explicit teaching of the digital tutee, whereas the other is more of an indirect measure of how well the student performs during gameplay when the tutee is learning through observation.

We computed a Pearson correlation coefficient (Table 7) for the two performance measures as well as for the two measures from the chat on students’ positive or negative attitudes towards the digital tutee and the feedback it provided: frequency of positive responses to feedback (freqPosAFB) and frequency of positive comments on the tutee’s intelligence and competence (freqPosCI). The comments on the tutees’ attitude had to be excluded due to the sparse data.

**Table 7.** The Pearson product-moment correlation coefficients for performance measures: answers to multiple choice questions, goodness, and chat measures: frequency of positive responses to feedback, and frequency of positive comments on the tutees’ intelligence and competence.

	1	2	3	4
1. Answers to multiple choice questions	-			
2. Goodness	0.338 **	-		
3. Pos. feedback responses	0.199	0.031		
4. Pos. comments on intelligence and competence	0.248 *	−0.016	0.593 **	-

\* Correlation is significant at the 0.05 level (2-tailed); \*\* Correlation is significant at the 0.01 level (2-tailed).

Overall, we found a significant correlation with large effect size ( $r(78) = 0.593, p < 0.01$ ) between the frequency of students’ positive answers to the digital tutee’s feedback and the digital tutees comments on its own intelligence and competence. There was a significant correlation of medium effect size ( $r(88) = 0.388, p < 0.01$ ) between the two performance measures; correctly answered multiple-choice questions and goodness (i.e., choosing the best card). We also found a significant correlation of small

effect size between how well the student answered the multiple-choice questions and the frequency of providing positive comments on the digital tutees' intelligence or competence ( $r(79) = 0.248, p < 0.05$ ).

When looking at the different student groups we found no significant correlation between performance measures for students with low self-efficacy. But we did find a significant correlation of medium effect size ( $r(39) = 0.359, p < 0.05$ ) between their proportion of correct answers to multiple-choice questions and the frequency of positive comments they provided on the digital tutees' intelligence or competence. The correlation between the frequency of positive comments they provided on the digital tutees' intelligence or competence and the frequency of positive feedback responses was also significant and of large effect size ( $r(39) = 0.698, p < 0.01$ ) (see Table 8).

**Table 8.** The Pearson product-moment correlation coefficients for students with low self-efficacy.

	1	2	3	4
1. Answers to multiple choice questions	-			
2. Goodness	0.204	-		
3. Pos. feedback responses	0.280	-0.130		
4. Pos. comments on intelligence and competence	0.359 *	0.130	0.698 **	-

\* Correlation is significant at the 0.05 level (2-tailed); \*\* Correlation is significant at the 0.01 level (2-tailed).

For the students with high self-efficacy another pattern came forth, see Table 9. There was a significant correlation of medium effect size ( $r(44) = 0.368, p < 0.05$ ) between the proportion of correctly answered questions and goodness, as well as a significant correlation between the frequency of positive responses to the feedback and the frequency of positive comments they gave on the digital tutees intelligence or competence ( $r(39) = 0.464, p < 0.01$ ). However, no significant correlation was found between their proportion of correct answers to the multiple-choice questions and the frequency of positive comments they provided on the digital tutee's intelligence or competence.

**Table 9.** The Pearson product-moment correlation coefficients for students with high self-efficacy.

	1	2	3	4
1. Answers to multiple choice questions	-			
2. Goodness	0.368 *	-		
3. Pos. feedback responses	0.040	0.041		
4. Pos. comments on intelligence and competence	0.113	-0.216	0.464 **	-

\* Correlation is significant at the 0.05 level (2-tailed); \*\* Correlation is significant at the 0.01 level (2-tailed).

This can be interpreted as students with low self-efficacy who express a more positive attitude towards their digital tutee in the sense of providing more positive comments on their digital tutee's intelligence and competence also perform better when they answer the digital tutee's multiple-choice question. However, they do not play better, with reference to how they choose cards (i.e., the goodness value). The competence of students with high self-efficacy seems to be the driving force in how they perform. In this group, students who play the game well choose good cards (having a high goodness value) also answer the multiple-choice questions better, regardless of their attitude towards their digital tutee as expressed through their chat comments on the tutee's intelligence and competence.

However, the differences in correlation between goodness and answers to multiple choice questions between the students with high and low self-efficacy, was not significant,  $Z = -0.816, p = 0.42$ , and neither was the correlation between positive comments on competence and intelligence and answer to multiple choice questions,  $Z = 1.11, p = 0.27$ . So, these results should be taken more as an indication of something that may be interesting to investigate further.



#### 4. Discussion

In [37] we drew the tentative conclusion that designing a digital tutee with low self-efficacy would be a good choice since the results suggested that students with low self-efficacy benefitted from interacting with a digital tutee with low rather than high self-efficacy. At the same time, students with high self-efficacy were not negatively affected when interacting with a digital tutee with low self-efficacy; rather they performed equally well when interacting with both types of tutees.

With the follow-up analysis carried out in this paper, we hoped to get a deeper understanding of these results by analyzing the interaction between student and his/her tutee. By looking at the dialogue we hoped to find interaction patterns that could explain why a digital tutee with low self-efficacy would be a better choice when designing pedagogical agents for educational software. The analysis was based on the dialogues between student and tutee that appeared as a chat after a finished game. It was in this chat the digital tutee expressed its self-efficacy (either high or low) when conversing/talking/interacting with the student.

Our findings can be summarized as follows:

- (i) Students responded more frequently to feedback from a digital tutee with low self-efficacy, and these responses were mostly positive.
- (ii) Students gave a digital tutee with low-self efficacy more positive comments about its intelligence and competence than they did to a digital tutee with high self-efficacy.
- (iii) Students' comments about the tutees attitude were almost exclusively given to the tutee with low self-efficacy. Most comments were positive, expressing that the tutee and/or student was doing well or were of an encouraging type. There were, however, also some comments that expressed frustration regarding the tutee's low opinion of itself.
- (iv) Students with low self-efficacy who expressed a more positive attitude towards their digital tutee, in the sense of providing more positive comments on the digital tutees intelligence and competence, also performed better when they answered the digital tutee's multiple-choice questions. However, they did not play better in the sense of choosing more appropriate cards (i.e., goodness value). For students with high self-efficacy we found another pattern, namely a relation between how well they played and how well they answered the questions asked by the digital tutee.

Below, we discuss how these findings can be understood in the light of the following three constructs: the protégé-effect, role modelling, and the importance of social relations. We know from previous studies that the protégé-effect is one of the underlying factors that make students who teach someone else (for example a digital tutee) learn more and be more motivated compared to students who learn for themselves [30]. That is, having someone who is dependent on you to learn and that you have responsibility for seems to lead to an increased effort. From our analysis, we see that students responded more frequently and more positively (finding (i)) to a tutee with low self-efficacy than to one with high self-efficacy. Many students tried to encourage a tutee with low self-efficacy when they for example commented on its attitude, saying things like "Tell yourself that you can win" or "Don't worry, you can do it!" The tutee with low self-efficacy also received fewer negative comments on its intelligence and competence than the tutee with high self-efficacy (finding (ii)).

Possibly, students treat a digital tutee with low self-efficacy in a more positive manner, in that they respond more frequently and more positively to its comments since such a tutee comes across as someone more in need of help and who is more subordinate compared to a digital tutee with high self-efficacy. The experience of having a protégé to care for and to support might be especially relevant for students with low self-efficacy, in this case, notably low self-efficacy in mathematic. These students will, more often than students with high self-efficacy in math, lack the experience of being someone who teaches someone else. Students with high self-efficacy are more likely to already in regular classes have taken a teacher role and assisted or supported a less knowledgeable and/or less confident peer.

Based on the findings of [41], that a person's self-efficacy may be influenced by observing someone else perform a special task, one could have suspected that a digital tutee with high self-efficacy would function as a role model and thus boost the students' self-efficacy. Seeing someone else doing something may boost the thought: "if (s)he can do it so can I". But in our analyses, we only found three instances of comments where the students agreed with the digital tutee when it expressed high self-efficacy, saying for example "It's good that you are confident, it will go well" and two of these comments came from students who themselves had high self-efficacy. Instead, a kind of reversed role modelling may be going on in which the student can be a role model for the tutee with low self-efficacy. In our analysis, we found that when the digital tutee expressed a very negative attitude some of the students were positive and encouraged it with wordings such as "You shouldn't be worried, you will make it" or "I know ... but you are pretty good, you just have to think positive" (finding (iii)). We did not find any comments where the student agreed with the tutee with low self-efficacy and expressed his or her own low self-efficacy. One can speculate that this is a result of the reversed roles that comes with teachable agents and learning by teaching, where the student takes the role as the teacher and the agent acts as the tutee.

Finally, we turn to the importance of the social relation with the tutee. Sjöden and colleagues [31] have previously shown that the presence of a digital tutee can have a positive impact on low-performing students. Students with low self-efficacy do not equal low-performing students, but there is often a correlation between the groups. Looking at our analysis, we found that for students with high self-efficacy, their performance (i.e., goodness value) correlated with how well they answered the digital tutees multiple-choice questions. This correlation is not surprising since someone who answers the questions correctly is likely to be good at choosing good cards. What is interesting however is that we did not find this correlation for students with low self-efficacy. Instead, we found a correlation between how well they answered the digital tutees multiple-choice questions and to what extent they gave positive comments regarding their tutees' competence and intelligence. It is possible that when students formed a social relationship with their digital tutees in the chat, by encouraging it, it had an effect on the students' engagement and performance in the interaction with the agent outside the chat. If the student feels that he or she is the more knowledgeable one, it might fuel the will to invest more in the learning since they act as the role model. For students with low self-efficacy this is presumably a situation they are not as used to as students with high self-efficacy, and they may therefore react more strongly to it.

### *Limitations*

Even though the digital tutee had the ability to talk about a wide array of topics its abilities were limited. You could sense that some of the students were a bit frustrated at points when the digital tutee could not answer the questions asked by the student. Maybe this led to more negative comments and frustration than otherwise would be the case.

The research questions focused on looking at agents and students on the extreme ends of the self-efficacy scale. The digital tutee was designed to have either clearly high or low self-efficacy and the analysis was restricted to the students with the highest and lowest self-efficacy score, with mid self-efficacy students excluded. Another way to perform the study would be to look at self-efficacy as a continuous metric and investigate if there are linear relationships between the students' self-efficacy and other variables. The choice to not do so was partly based on limitations in resources to code the chat dialogues.

## 5. Conclusions

In this paper, we have explored how a digital tutee's self-efficacy might affect how students interact with it, given that the students themselves have high or low self-efficacy. A tentative conclusion from our previous paper [37] was that it is more beneficial to design a digital tutee with low self-efficacy than one with high self-efficacy. Our present analysis shows that this, all in all, is a good guideline, and that the underlying reason can be that a digital tutee with low self-efficacy boosts the protégé effect more, and also promotes a reversed role modelling where the student can boost herself through boosting the digital tutee. However, we have seen some indications that some students can find a digital tutee with low self-efficacy frustrating. In some cases, this occurs when the students or their digital tutee performed well but the tutee expressed a negative attitude. Thus, it may be that the tutee's self-efficacy needs to be more adaptive and better reflect the rate at which it actually learns, which in turn reflects the proficiency of the student that is teaching it.

There may also be differences within the group of students with low self-efficacy. Students with low self-efficacy tend to overlap with the group of low-performing students, and during our classroom visits we can usually identify a subgroup of students that do not perform well due to not caring or trying, while we see others who do care and try, but fail nevertheless. In this study it was the subgroup of students with low self-efficacy that engaged with the agent in the chat, that also performed better in situations where the agent asked questions, which could be an indication of differences within this group of students.

**Supplementary Materials:** Chatlogs and statistical data are available online at <http://www.mdpi.com/2227-7102/9/3/197/s1>.

**Author Contributions:** The first author designed and carried out the data collection. Design and analysis of the study were performed by the first and second author. The statistical analysis was done by the second author. The paper was jointly written and edited by both authors.

**Funding:** This research received no external funding.

**Acknowledgments:** Thanks to Agneta Gulz, Magnus Haake and Nils Dahlbäck for helpful comments on previous draft of this paper. A thank you also to the four coders Maximilian Roszko, Kristian Månsson, Eva-Maria Ternblad and Erika Rönningberg. Also, a thank you to all the teachers and students who participated in the study.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A

Nr: \_\_\_\_\_

I would like you to try to estimate how *well you would do* if you were asked to solve a number of tasks. You do NOT have to solve the tasks. Just mark how good you *would be* at solving them.

**How good would you be at solving these tasks?**

	Really Bad	Bad	Neither nor	Good	Really Good
1. $1136 + 346$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. $184 - 64$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. What number is missing? $670 - \underline{\quad} = 485$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. You have the number 274, will the result end with 00 if you add 3826?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Which of the totals is largest $295 + 16 + 1719$ or $32 + 2234 + 123$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. What number do you get if you swap the hundred and the ten in 437?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. What number has the largest value in 6275?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

---

## References

- Graesser, A.C.; Wiemer-Hastings, K.; Wiemer-Hastings, P.; Kreuz, R. Tutoring Research Group. AutoTutor: A simulation of a human tutor. *Cogn. Syst. Res.* **1999**, *1*, 35–51. [[CrossRef](#)]
- Chan, T.W.; Chou, C.Y. Exploring the design of computer supports for reciprocal tutoring. *Int. J. Artif. Intell. Educ.* **1997**, *8*, 1–29.
- Hietala, P.; Niemirepo, T. The competence of learning companion agents. *Int. J. Artif. Intell. Educ.* **1998**, *9*, 178–192.
- Kim, Y.; Baylor, A.L.; PALS Group. Pedagogical agents as learning companions: The role of agent competency and type of interaction. *Educ. Technol. Res. Dev.* **2006**, *54*, 223–243. [[CrossRef](#)]
- Johnson, W.L.; Rickel, J.W.; Lester, J.C. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *Int. J. Artif. Intell. Educ.* **2000**, *11*, 47–78.
- Graesser, A.C.; Person, N.; Harter, D. Tutoring Research Group. Teaching tactics and dialog in AutoTutor. *Int. J. Artif. Intell. Educ.* **2001**, *12*, 257–279.
- Baylor, A. Beyond butlers: Intelligent agents as mentors. *J. Educ. Comput. Res.* **2000**, *22*, 373–382. [[CrossRef](#)]
- Baylor, A.L.; Kim, Y. Simulating instructional roles through pedagogical agents. *Int. J. Artif. Intell. Educ.* **2005**, *15*, 95–115.
- Blair, K.; Schwartz, D.; Biswas, G.; Leelawong, K. Pedagogical agents for learning by teaching: Teachable agents. *Educ. Technol. Sci.* **2007**, *47*, 56–61.
- Bargh, J.A.; Schul, Y. On the cognitive benefits of teaching. *J. Educ. Psychol.* **1980**, *72*, 593–604. [[CrossRef](#)]

11. Annis, L.F. The processes and effects of peer tutoring. *J. Educ. Psychol.* **1983**, *2*, 39–47.
12. Renkl, A. Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learn. Instr.* **1995**, *5*, 21–36. [\[CrossRef\]](#)
13. Atkinson, R.K. Optimizing learning from examples using animated pedagogical agents. *J. Educ. Psychol.* **2002**, *94*, 416–427. [\[CrossRef\]](#)
14. Baylor, A.L. Expanding preservice teachers' metacognitive awareness of instructional planning through pedagogical agents. *Educ. Technol. Res. Dev.* **2002**, *50*, 5–22. [\[CrossRef\]](#)
15. Moreno, R.; Mayer, R.E.; Spires, H.A.; Lester, J.C. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cogn. Instr.* **2001**, *19*, 177–213. [\[CrossRef\]](#)
16. Kim, Y.; Baylor, A. Pedagogical agents as social models to influence learner attitudes. *Educ. Technol.* **2007**, *47*, 23–28.
17. Kim, Y.; Hamilton, E.R.; Zheng, J.; Baylor, A.L. Scaffolding learner motivation through a virtual peer. In Proceedings of the 7th International Conference on Learning Sciences, Bloomington, IN, USA, 27 June–1 July 2006; International Society of the Learning Sciences: Bloomington, IN, USA; pp. 335–341.
18. Kim, Y.; Wei, Q.; Xu, B.; Ko, Y.; Ilieva, V. MathGirls: Toward developing girls' positive attitude and self-efficacy through pedagogical agents. In Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED), Los Angeles, CA, USA, 9 July–13 July 2007; IOS Press: Los Angeles, CA, USA, 2007; pp. 119–126.
19. Pareto, L.; Arvemo, T.; Dahl, Y.; Haake, M.; Gulz, A. A teachable-agent arithmetic game's effects on mathematics understanding, attitude and self-efficacy. In Proceedings of the International Conference on Artificial Intelligence in Education, Auckland, New Zealand, 28 June–2 July 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 247–255.
20. Kim, Y.; Lim, J.H. Gendered socialization with an embodied agent: Creating a social and affable mathematics learning environment for middle-grade females. *J. Educ. Psychol.* **2013**, *105*, 1164–1174. [\[CrossRef\]](#)
21. Kim, Y.; Baylor, A.L.; Shen, E. Pedagogical agents as learning companions: The impact of agent emotion and gender. *J. Comput. Assist. Lear.* **2007**, *23*, 220–234. [\[CrossRef\]](#)
22. Gulz, A.; Haake, M. Design of animated pedagogical agents—A look at their look. *Int. J. Hum Comput. Stud.* **2006**, *64*, 322–339. [\[CrossRef\]](#)
23. Baylor, A.L. Promoting motivation with virtual agents and avatars: Role of visual presence and appearance. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2009**, *364*, 3559–3565. [\[CrossRef\]](#)
24. Kim, Y. The role of agent age and gender for middle-grade girls. *Comput. Sch.* **2016**, *33*, 59–70. [\[CrossRef\]](#)
25. Wang, N.; Johnson, W.L.; Mayer, R.E.; Rizzo, P.; Shaw, E.; Collins, H. The politeness effect: Pedagogical agents and learning outcomes. *Int. J. Hum Comput. Stud.* **2008**, *66*, 98–112. [\[CrossRef\]](#)
26. Veletsianos, G. The impact and implications of virtual character expressiveness on learning and agent learner interactions. *J. Comput. Assist. Lear.* **2009**, *25*, 345–357. [\[CrossRef\]](#)
27. Uresti, J.A.R. Should I teach my computer peer? Some issues in teaching a learning companion. In Proceedings of the International Conference on Intelligent Tutoring Systems, Montréal, QC, Canada, 19–23 June 2000; Springer: Berlin/Heidelberg, Germany, 2000; pp. 103–112.
28. Kim, Y. Desirable characteristics of learning companions. *Int. J. Artif. Intell. Educ.* **2007**, *17*, 371–388.
29. Okita, S.Y.; Schwartz, D.L. Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *J. Learn. Sci.* **2013**, *22*, 375–412. [\[CrossRef\]](#)
30. Chase, C.C.; Chin, D.B.; Oppezzo, M.A.; Schwartz, D.L. Teachable agents and the protégé effect: Increasing the effort towards learning. *J. Sci. Educ. Technol.* **2009**, *18*, 334–352. [\[CrossRef\]](#)
31. Sjöden, B.; Tärning, B.; Pareto, L.; Gulz, A. Transferring teaching to testing – an unexplored aspect of teachable agents. In Proceedings of the International Conference on Artificial Intelligence in Education, Auckland, New Zealand, 28 June–2 July 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 337–344.
32. Bickmore, T.; Cassell, J. Small talk and conversational storytelling in embodied conversational interface agents. In *Papers from the AAAI Fall Symposium on Narrative Intelligence*, Cape Cod, MA; AAAI Press: Menlo Park, CA, USA, 1999; pp. 87–92.
33. Cassell, J.; Bickmore, T. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Model. User Adap.* **2003**, *13*, 89–132. [\[CrossRef\]](#)

34. Veletsianos, G. How do learners respond to pedagogical agents that deliver social-oriented non-task messages? Impact on student learning, perceptions, and experiences. *Comput. Hum. Behav.* **2012**, *28*, 275–283. [[CrossRef](#)]
35. Tärning, B.; Haake, M.; Gulz, A. Off-task Engagement in a Teachable Agent based Math Game. In Proceedings of the 19th International Conference on Computers in Education (ICCE-2011), Chiang Mai, Thailand, 28 November–2 December 2011; Hirashima, T., Ed.; Asia-Pacific Society for Computers in Education: Chiang Mai, Thailand, 2011.
36. Silvervarg, A.; Haake, M.; Pareto, L.; Tärning, B.; Gulz, A. Pedagogical agents: Pedagogical interventions via integration of task-oriented and socially oriented conversation. Paper presented at The Annual Meeting of the American Educational Research Association, New Orleans, LA, USA, 8–11 April 2011; pp. 1–16.
37. Tärning, B.; Silvervarg, A.; Gulz, A.; Haake, M. Instructing a Teachable Agent with Low or High Self-Efficacy—Does Similarity Attract. *Int. J. Artif. Intell. Educ.* **2019**, *29*, 89–121. [[CrossRef](#)]
38. Pareto, L. A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *Int. J. Artif. Intell. Educ.* **2014**, *24*, 251–283. [[CrossRef](#)]
39. Bandura, A.; Barbaranelli, C.; Caprara, G.V.; Pastorelli, C. Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Dev.* **1996**, *67*, 1206–1222. [[CrossRef](#)] [[PubMed](#)]
40. Silvervarg, A.; Haake, M.; Gulz, A. Educational potentials in visually androgynous pedagogical agents. In Proceedings of the International Conference on Artificial Intelligence in Education, Memphis, TN, USA, 9–13 July 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 599–602.
41. Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* **1977**, *84*, 191–215. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).