Using Iterative Design to Create Efficacy-Building Social Experiences With a Teachable Robot

Nichola Lubold, Arizona State University, nichola.lubold@asu.edu
Erin Walker, Arizona State University, erin.a.walker@asu.edu
Heather Pon-Barry, Mount Holyoke College, ponbarry@mtholyoke.edu
Yuliana Flores, University of Washington, yulif21@uw.edu
Amy Ogan, Carnegie Mellon University, aeo@cs.cmu.edu

Abstract: Teachable robots are a form of social robot for education, where learners engage in conversation to teach the robot like they would a peer. Part of the popularity of social robots is their ability to utilize social channels of communication to foster productive social experiences, interactions which help individuals grow and develop. Teachable robots have potential to utilize social channels of communication to create social experiences which can help learners develop self-efficacy, an individual's belief in their ability to succeed. In this paper, we present a fully autonomous robot for middle school math; we iterate through three design phases and analyze responses to identify how to better foster productive social experiences for self-efficacy. We report six design recommendations; for example, for low self-efficacy individuals, an ideal design should incorporate problem-solving statements and positivity to foster social experiences of mastery and social persuasion.

Introduction

In this project, we investigate how robotic teachable agents can be designed to create social self-efficacy experiences, or social interactions that may aid learners in improving their self-efficacy, their belief that they can succeed in a domain. Teachable agents are agents that learners can teach about a subject domain (Chou, Chan, and Lin, 2003). When teaching others, learners attend more to the problem, reflect on misconceptions when correcting their peers' errors, and elaborate on their knowledge as they construct explanations (Roscoe and Chi, 2007). Learning by teaching can improve domain knowledge (Kauchak and Eggen, 1993), self-efficacy (Frager and Stern, 1970), and peer attitudes (Griffin and Griffin, 1998).

Implementing a teachable agent as a physical robot might enhance the learner's social experience, which may then improve learning and motivation within a domain (Dillenbourg et al., 2016). Social factors play a strong role when learners teach agents. Those that feel invested in their agents' learning (Leelawong and Biswas, 2008) or feel rapport with their agent (Ogan et al., 2012) are more likely to benefit. Because robots have a physical presence and rich channels of communication, they have under some circumstances developed a deeper social engagement with users than their virtual equivalent (Liu et al., 2013). The use of teachable robots may increase users' social engagement and thus enhance the impact of social factors in learning by teaching.

Our goal is to implement a teachable robot to improve learners' self-efficacy. Self-efficacy, or belief in one's ability to succeed, is a significant predictor of task performance (Bandura and Schunk, 1981). Self-efficacy in STEM plays a major role in learner persistence and success. However, it is not easy for learners to develop self-efficacy. We aim to provide a learner with the following social experiences, theorized by Bandura (1977) to have positive effects on self-efficacy:

1 **Mastery experiences**, facilitated by experiences where learners successfully teach the robot
2 **Vicarious learning experiences**, where a social model exemplifies good learning practices
3 **Social persuasion**, through social interaction learners are convinced of their success as domain experts

Bandura further hypothesizes that a learner's rapport, or sense of connection, with a collaborating partner, might enhance the effects of these pathways. This suggests a fourth goal for our robot:

4 **Building rapport**, or a feeling of connection, with the learner

It is not clear how exactly a robot can be designed to achieve these goals. In this project, we examine the design of social experiences within the context of a teachable robot for middle school mathematics, where learners teach the robot about ratio problems. Our contributions are two-fold. First, we present the implementation of a fully autonomous teachable robot for mathematics, Nico. Nico is a Nao robot that can interact with the learner using spoken dialogue and realistic gestures. The richness of Nico's interactions makes it a unique platform for exploring social factors and effects on self-efficacy. Second, we present the results of a multi-phase iterative design process with 14 learners; through this process, we specifically explore how Nico's dialogue can foster these social, self-efficacy experiences. We pose the following research questions:

- How do different dialogue design strategies based on human-human peer tutoring and theories of rapport enhance mastery, vicarious experience, social persuasion, and rapport with a teachable robot?
- How might individual differences, such as initial self-efficacy, influence responses to different dialogue design strategies in a teachable robot?

Motivated by the literature on peer tutoring and rapport, we iterated on the design of Nico's dialogue for each social experience. We explored how Nico's dialogue can facilitate **mastery** by balancing the challenge learners face in articulating knowledge while enabling them to feel successful. We investigated how Nico's dialogue can model different learning practices such as question-asking and optimism and spark a corresponding response in the learner as **vicarious experience**. We explored both subtle and overt approaches to **social persuasion**. Finally, we explored **rapport** by iterating over rapport signaling behaviors of friends versus strangers. We analyzed learners' responses in each phase, yielding six design recommendations with an emphasis on how different individuals might benefit from different dialogue designs.

Prior work

Human-human peer tutoring literature indicates the importance of social factors and the presence of social outcomes in tutoring interactions. One meta-analysis suggests that peers who give help experience many social and motivational benefits, including improved attitudes towards school, academic self-concept, and academic self-efficacy (Robinson et al., 2005). Recent work on peer tutoring has suggested the rapport between two collaborating partners may correlate with learning outcomes (Ogan et al., 2012b). Madaio, Ogan, and Cassell (2016) found when tutees and tutors are friends, they engage in different tutoring strategies which may be more productive. These factors, if reproduced, may contribute to productive social experiences with a teachable robot.

Our work is also inspired by the broader teachable agent literature, where learners teach a virtual agent. Teaching a computer agent can be beneficial for the learner doing the teaching: it can lead to more learning than being taught by a computer agent (Leelawong and Biswas, 2008) and can be more effective than classroom instruction at improving learning and self-efficacy (Pareto et al., 2011). Chase and colleagues theorized that these benefits originate from the protégé effect, which occurs when learners become more motivated to teach their agent because: (1) they feel more responsible for their agent's success and, (2) failures reflect on the agent rather than on them (Chase et al. 2009), and this may then facilitate learners experiencing greater mastery and social persuasion. Other human-agent work has suggested that the benefits of teachable agents come from feeling more rapport for an agent partner (Ogan et al., 2012a). Less work has focused on explicitly designing interactions to promote social experiences. One exception is Gulz, Haake, and Silvervaarg (2011) who demonstrate that adding off-task dialogue to a teachable agent can facilitate learning and positive outcomes.

Despite the promise of teachable agents in general, there has only been a small body of prior work on teachable robots. Co-Writer is a Nao robot that learners teach about handwriting. Through adaptation of the robot's learning behavior, studies have shown the robot can engage learners in the task and potentially promote motivation and self-confidence (Jacq et al., 2016). Tanaka and Matsuzoe explored a Nao robot that learners can teach about vocabulary through physical demonstration (Tanaka and Matsuzoe, 2012). The interaction showed teaching a robot can support learning and suggested that both verbal and gestural communication is natural for teaching robots. rTAG is a Lego Mindstorms-based learning environment where learners teach the robot coordinate geometry (Walker et al., 2016). rTAG explored how physical embodiment affects social engagement; social engagement increased in low prior knowledge learners but learning gains decreased. A subsequent iteration of rTAG known as Quinn explored the role of dialogue; Lubold and colleagues found a combination of verbal and paraverbal social dialogue influenced perceptions of social presence (Lubold, Walker, and Pon-Barry, 2016). Collectively, this work indicates that teachable robots have potential to support learning, but the role of social factors remains unclear. In rTAG, physical embodiment alone was not enough to foster social engagement, while mastery experiences in Co-Writer was. Quinn revealed verbal and paraverbal features may enhance social experience but how dialogue can be designed to promote self-efficacy is an open question. We build on this prior work by introducing a fully autonomous teachable robot and exploring how empirically motivated, theoretically grounded dialogue can be designed to foster social experiences and self-efficacy.

System

We have developed an autonomous teachable Nao robot for middle school mathematics named Nico. In addition to the Nao robot, our system includes a user interface (UI) and a dialogue system. The learning domain for the teachable robot is middle school mathematics, specifically reasoning about ratios. We designed four narrative-style ratio word problems for learners to teach Nico based on the Common Core Standards for 6th grade. For each problem, Nico and the learner are given partial information; Nico requests the learner's help in how to use ratios to solve for the missing information. An example problem is depicted in Figure 1.

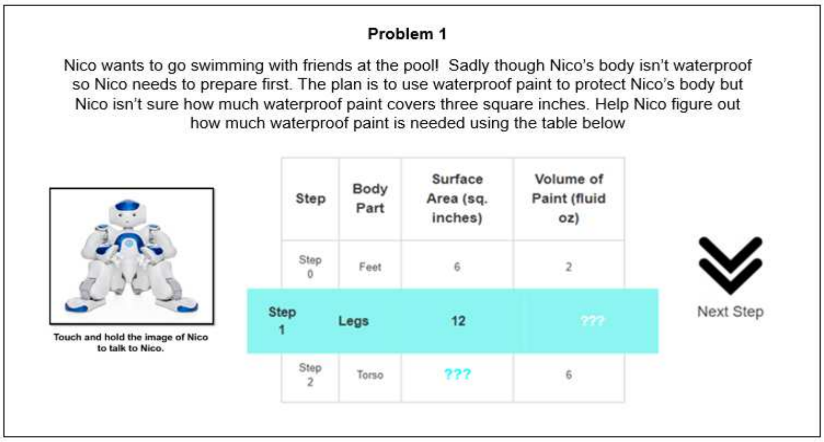

Figure 1. Example of a problem as depicted in the user interface.

Learners interact with Nico using spoken language and a touch-screen interface on a tablet computer (Microsoft Surface Pro) that displays each problem. The UI supports speech recognition and displays visual progress as the learner moves through the problems. To speak to Nico, the learner presses a button on the interface. After they are finished speaking, a notice appears on the interface indicating that Nico is 'thinking' while the system processes the input and generates a response. Average response time is less than four seconds. The UI tracks progress as the learner guides Nico through each problem step at their own pace, using buttons to advance forward. The current step is highlighted and enlarged on the screen. When Nico 'answers' a step, the corresponding table cell is updated from question marks (see Figure 1) to the correct answer.

The dialogue system is composed of an automatic speech recognizer, dialogue manager, and the Nao robot's text-to-speech synthesizer. The interface captures the user's speech using the tablet's default microphone and speech recognition is performed using the Google Speech API. We find an average word error rate of 22.2%. Nico takes the initiative in the dialogue whenever a learner starts a problem, typically to request help. We utilize the Artificial Intelligence Markup Language (AIML) (Wallace, 2003) to process the learner's input and identify Nico's response. AIML is an XML-compliant pattern matching language. We designed and developed the dialogue by mapping all possible solutions for each problem step to potential explanations. After a learner speaks, a potential explanation pathway is identified based on keywords in the speaker's speech. If a potential explanation cannot be matched, Nico requests a clarification. For example, Nico might say "I'm actually not sure what you said. Can you try saying it again but differently?" Clarification requests occurred on average 4% of the time. The frequency of clarification requests did not differ substantially across design phases.

In addition, we programmed Nico to produce eight emblematic or easily recognizable gestures. Gestures include waving 'hello,' nodding head as in 'yes,' shaking head 'no,' putting hands on hip to make a point, raising either hand, raising hands in celebration, and shrugging. The dialogue manager identifies whether there is an appropriate gesture in the set to accompany Nico's utterance. We also enabled 'autonomous life', a default capability of the Nao robot which displays slight movement and listening behavior. Gesture and autonomous life were present in all phases. We do not explore gesture in the iterative design in this work, rather focusing on dialogue. The role of gesture in social, self-efficacy experiences is an opportunity for future work.

Iterative design

With the platform described above, we can iterate over the design of a teachable robot's dialogue and explore (1) how different dialogue design strategies might enhance mastery, vicarious experience, social persuasion, and rapport, and (2) the role of individual differences in response to different strategies.

Method

Fourteen participants between ages 10 and 13 (M: 11, SD: 1.0, 4 female/10 male) participated across three exploratory design iterations with 5 participants (1 female/4 male) in the first phase, 5 (2 female/3 male) participants in the second phase, and 4 participants (1 female/3 male) in the final phase. All participants were native English speakers. Sessions lasted 90 minutes. Participants were recruited via flyers and emails shared during local programs offered to middle schoolers on the university campus.

The procedure for each session was the same across all phases. Participants began by completing a 10-minute pretest on ratios. Next, each participant was given a pre-survey on self-efficacy and comfort-level towards robots. Before interacting with Nico, participants were given a few minutes to review the worked-out solutions to the problems they were to teach. Participants then watched a 3-minute video introducing Nico. After the video, Nico initiated a brief 'introduction' interaction by saying "Hello, it's nice to meet you. What is your name?" The 'introduction' gave participants an opportunity to practice talking to Nico before teaching. Participants utilized the tablet and spoken dialogue to teach Nico and moved through the problems at their own pace. Time to complete teaching the problems varied from 12 to 35 minutes. After the interaction, participants completed a post-survey; twelve participants also took a 10-minute posttest (isomorphic to the pretest). We then conducted structured interviews; the same questions were asked of every participant.

To evaluate the design and impact of each phase, we collected self-reported measures of rapport, self-efficacy, and learning gains and performed a comparative analysis on the interviews, experimenter observations, and dialogue transcripts. For the self-reported rapport, we posed a set of 14 Likert scale questions about rapport to each participant. The questions were based on a combination of prior work exploring rapport in human-human (Tickle-Degnen and Rosenthal, 1990), human-agent (Cassell, Gill, and Tepper, 2007), and human-robot interactions (Lubold, Walker, and Pon-Barry, 2016). Questions related to feelings of general rapport (i.e., "Nico and I worked well together"), positivity (i.e., "I liked Nico, Niko liked me"), attention ("Nico paid attention to me"), and coordination ("I was awkward in talking to Nico"). We asked participants in post-interviews to explain their understanding and interpretation of each survey question; these validations enabled small iterative changes to improve the wording. We aggregated the questions for each participant into an average rapport score.

Self-efficacy towards math was measured with six questions based on work by the Friday Institute for Technology (2008). Participants answered the six questions both before and after interacting with Nico. Averaging the responses, we calculated whether participants experienced a change in self-efficacy as post-score minus pre-score. Additionally, we asked how comfortable individuals were interacting with robots and human-looking robots. Finally, we calculated learning from the pre- and post-tests as a normalized learning gain score, as recommended in (Hake, 2002).

We focused our qualitative analyses on the interviews, experimenter observations, and transcripts of the interaction dialogue. Since we are interested in identifying the degree to which the dialogue can influence social experience, we coded the data for themes regarding mastery experience, vicarious experience, social persuasion, and rapport. We identified a set of decision rules for identifying themes, as suggested by Miles, Huberman, and Saldana (1994). For example, for mastery to be present, the learner must give evidence they feel Nico learned. An example decision rule for mastery was: participant is marked as having felt a degree of mastery based on the presence of either (a) did the learner give any reference to Nico having 'learned' or (b) did the learner mention any observations regarding Nico's learning (i.e. "Nico said 'I understand' so he was learning"). Two researchers compared and contrasted identified rules, themes and analyses to validate consistency; findings were triangulated across all data to ensure all pieces were mutually supportive.

Phase I

In Phase I, we began with dialogue which might foster mastery, vicarious experience, social persuasion and rapport based on human-human peer tutoring. To foster **mastery**, Nico challenged participants to explain each step thoroughly before reaching an understanding of the solution. To foster **vicarious experience** and **social persuasion**, we introduced a question-answer design based on human-human peer tutoring (Graesser and Person, 1994) which found that tutees typically ask questions with low specificity. We designed Nico's dialogue as approximately 80% questions with a tendency towards low specificity. In addition, Nico was responsive, giving a sense of paying attention by acknowledging answers to questions. To build **rapport**, we based the dialogue design on behaviors suitable to strangers in peer tutoring with an interest in whether it is better for Nico's dialogue to model that of a stranger or a friend. In human-human peer tutoring, tutees who are strangers ask more questions and these questions tend to be shallow (Madaio, Ogan, and Cassell, 2016); tutors and tutees who are strangers are more polite (Ogan et al. 2012b). Nico's dialogue already included questions; we further designed these to be shallow and more polite. Table 1 gives an example of the dialogue.

Table 1: Example dialogue from each phase. In Phase I, Nico asks questions, is polite but slow to understand, and has low specificity. Phase II maintains the same level of questions and politeness but Nico is more specific, understands more quickly, and exhibits positivity. Phase III maintains the positivity and specificity of Phase II but Nico is less polite, asks fewer questions, and is slightly slower to understand than Phase II

Phase One	Phase Two	Phase Three
User: Nico, first you're going to look at the ratio between the 6 and the 2. You can set it equal to 9. *Nico*: Okay but I'm not sure I get it. How do we do that? *User*: You put 6 over 2 equal to 9 over something. *Nico*: Oh okay like in a fraction? *User*: Yeah exactly *Nico*: I get it. We put 6 over 2 equal to 9. And then what do we do? *User*: Then you're gonna cross multiply 2 times 9. *Nico*: So then we multiply and then what do we do?	*User*: Nico, first you're going to look at the ratio between the 6 and the 2. You can set it equal to 9. *Nico*: Oh okay! Great! We put 6 over 2 like in a fraction? *User*: Yeah exactly and it is equal to 9 over something *Nico*: These problems are hard but I think I'm getting it, thanks to you! Now do we multiply 9 by 2? *User*: Yes so you're gonna cross-multiply to get 18 *Nico*: I get it; you're so helpful! What do we do after we have 18?	*User*: Nico, first you're going to look at the ratio between the 6 and the 2. You can set it equal to 9. *Nico*: So we put 6 over 2 like in a fraction, then we can use the ratio to figure it out. *User*: Yeah cause you have two ratios so you can multiply *Nico*: Oh okay! So we can multiply. *User*: Yes so you're gonna cross-multiply to get 18 *Nico*: Sweet! So then we have 9 times 2 is 18.

Five participants took part in this phase, 4 males and 1 female. The results for the rapport, change in self-efficacy and qualitative observations are summarized in Table 2. We found that for **mastery experience** four of five participants in this phase were not convinced of Nico's learning, reporting Nico only "kind of learned" (P1, P2). We observed no evidence of **vicarious experience.** Participants' did not appear to experience or observe Nico's model of question-asking and attention through responsiveness. The majority of the participants in this phase did not exhibit any form of **social persuasion**; they did not feel like Nico learned and also felt that they were not successful as tutors. The one participant who felt like Nico learned did not attribute Nico's success to himself. P5 felt that Nico succeeded despite his own flaws as a tutor. Finally, participants were largely neutral in the degree of **rapport** they felt for Nico. Few of the participants exhibited any sense of general rapport for Nico, and while participants responded in the interviews that while they 'liked' Nico (P1, P2, P3), Nico was still 'a robot, not a person' (P2, P3). Differences in verbal behaviors emerged; unlike the others, P5 praised Nico and was more inclusive.

Individuals overall did not appear to be having productive social experiences; we found no evidence of mastery, vicarious experience, social persuasion or rapport. We did find one participant with dissimilar responses, which suggests individual differences play a pertinent role in these types of experiences.

Phase II

In our first design phase, we attempted to facilitate experience of **mastery** by challenging learners to explain each step to Nico. However, they clearly did not experience mastery. It was possible the content of the problems and act of tutoring is challenge enough; for Phase II, we increased the speed at which Nico understands and we increased the level of specificity to see how this influences mastery. For **vicarious experience**, we kept the question-based design and responsiveness indicating attention but we explored whether other learning practices may be more suitable to vicarious experience with Nico. Positive social behavior during learning and staying optimistic in the face of challenging problems is correlated with learning outcomes (Pampaka, Williams, and Hutcheson, 2012). We introduced positivity (i.e. "Oh okay! Great!") and optimism (i.e. "these problems are hard but I think I'm getting it") into Nico's dialogue. For **social persuasion**, we had explored a subtle approach in Phase I; in Phase II, we introduced an overt design by framing messages to give outright encouragement success as a tutor (i.e. "You're so helpful!"). While these messages could be perceived as disingenuous, participants' belief in themselves may be positively influenced. Finally, for **rapport**, 'stranger-like' behaviors may have been distancing but it was also possible there were too few dimensions of behavior. We incorporated additional rapport-building behaviors while maintaining a model consistent with that of a stranger. Individuals who are strangers may introduce more positivity when building rapport (Tickle-Degnen and Rosenthal 1990). Increasing Nico's positivity to build rapport aligned with supporting vicarious experience and social persuasion.

Table 2: Summary of results and observations for each design phase; qualitative observations aggregate number of decision rules met for all participants in a particular phase for a given social experience

% of Decision Rules Met
0–20%
20–50%
50–75%
75–100%

	Mastery	Rapport	Vicarious Experience	Social Persuasion	Self-reported rapport	Δ in Math Self-Efficacy
Phase 1					3.52 (1.2)	.16 (.19)
Phase 2					3.91 (1.2)	.50 (.20)
Phase 3					4.67 (.29)	.68 (.17)

Five participants took part in this design phase, 2 females and 3 males. We found an increased number of participants (4 of 5) exhibited **mastery**, feeling Nico learned and that this learning was due to their tutoring. They noted that when Nico was specific, as in "Oh I guess we divide six by three?" (P7), they felt he was learning from what they told him. We still did not see any acknowledgement from the participants of **vicarious experience**; they did not comment on Nico's question-asking, attention, or positivity and optimism. In terms of s**ocial persuasion**, while most of the participants in this phase felt Nico had learned, few felt they were "good" tutors. It is possible we over-simplified the process for explaining steps to Nico, leading learners to feel Nico was just very smart, very intelligent, as one participant noted - "I didn't explain it very well but he was really smart so he got it". Also, P8 felt Nico's praise, designed to socially persuade the learners, was "undeserved." Finally, average **rapport** was higher in this phase. All five participants in post-interviews expressed higher engagement and two expressed feelings of accountability. These two learners also had the largest corresponding changes in self-efficacy, and praised Nico, "Good job" and "Nice job."

In this second design phase, evidence of productive social experiences increased. More participants expressed a sense of mastery, higher social engagement, and behaviors expressive of rapport. However, experiences of social persuasion in the form of feeling responsible for Nico's success and especially evidence of vicarious experiences of good learning practices were not substantial.

Phase III

In the second design phase, we observed positive responses to **mastery**. However, we also observed that individuals felt Nico was "very smart" and they did not experience **social persuasion**. The prior phase may have over-simplified the tutoring task in a way that could not be overcome by either subtle or overt persuasion and contributing to feelings that Nico's praise was disingenuous. We re-adjusted the level of challenge required in explaining to Nico how to solve the problems where Nico is slightly slower to understand than in phase II. We also did not see substantial evidence of **vicarious experience**; for this phase, we decided to focus on how increasing rapport might influence vicarious experience. For **rapport** in this final design phase, we emphasized behavior which would typically be found between friends rather than strangers. For example, Madaio, Ogan, and Cassell (2016) found that tutees who are friends tend to verbalize problem-solving statements more often than asking questions. We modified Nico's dialogue to incorporate statements about problem solving.

For this final phase, we had four participants take part, 3 male and 1 female. Again, we found evidence that participants experienced **mastery**, implying that slowing Nico's understanding did not influence whether participants felt like Nico learned. We did find evidence of **vicarious experience**. Participants commented on how teaching Nico was like teaching a friend and three participants noted the positivity of Nico's learning behavior. For example, P11 stated that he "doesn't get mad" and "Nico doesn't get frustrated at you" (P13), he stays "positive." Participants also noted Nico "doesn't get distracted as people tend to do" (P13) and was a good listener (P12). All four participants in this phase gave evidence of feeling **socially persuaded** that they taught Nico. It was "because they explained it well that he understood it" (P11, P12, P13). In post interviews, participants' comments reflected feelings of accountability in helping Nico learn as well. This phase had the highest **rapport** compared to the two previous design iterations; participants commented Nico "reminds me of my friend," is "pretty cool," "funny," and "cute." Verbal behaviors when interacting with Nico showed two out of the four participants praised Nico. We find their use of praise similar to prior phases. However, one participant gave Nico a little sass. (P13: "Thank you, Nico. Now get back to the questions!")

This iteration resulted in the most evidence of social experiences for enhancing self-efficacy. Participants continued to express a sense of mastery and high expressions of rapport. We finally saw evidence that individuals vicariously experienced models of good learning practices, and individuals not only felt success in the task but they expressed feelings of responsibility for that success.

Cross-phase trends

In addition to qualitative observations, we measured rapport, self-efficacy, and learning. The average rapport and change in self-efficacy are summarized for each design phase in Table 2. While participants in different design phases experienced different interactions, we explored cross-phase trends for insight into overall design directions. We found a significant correlation ($r = .71$, $p = .02$, $n = 10$) between rapport ($M = 4.0$, $SD = .9$) and learning gains ($M = .47$, $SD = .2$). We also found that rapport is significantly correlated with change in self-efficacy ($M = .46$, $SD = .3$) across all participants ($r = .62$, $p = .03$, $n = 14$). Change in self-efficacy was not correlated with learning gain ($r = .46$, $p = .17$, $n = 10$). These results support the theoretical argument that rapport is related to learning and self-efficacy and imply designing to enhance rapport may result in positive outcomes.

We did not observe a correlation between self-efficacy and learning. While this may be due to the iterative design and the small sample size, it is possible the relationship is obscured by individual differences. Within each phase we observed a single individual who was very socially engaged, from their self-reported rapport to their interview responses and verbal behaviors. Regardless of phase, these individuals praised Nico more, included Nico in the learning process with inclusive language, and were more likely to anthropomorphize Nico. Viewing Nico as socially and cognitively capable, these learners had high social responses, a low bar for social experience, and higher gains. Comparatively, individuals with the lowest rapport and the lowest change in self-efficacy (P3, P6, and P8) responded to Nico with less inclusive language, little to no praise, and spoke of Nico as "the robot." We found two other individuals, who interacted with Nico in phase III, reported initial self-efficacy scores as low as P3, P6, and P8; however, their change in self-efficacy was much higher. This suggests the third design phase may have been more effective for individuals with low self-efficacy.

Discussion and conclusion

In this paper, we described Nico, a fully autonomous teachable Nao robot for mathematics learning that can interact with learners using natural language. We explored through iterative design (1) how different dialogue design strategies can foster four social, self-efficacy experiences: mastery, vicarious learning, social persuasion, and rapport and (2) how individual differences influence responses to different dialogue design strategies. Our final design, which yielded the highest self-reported feelings of self-efficacy and rapport, was the most successful at fostering these four experiences. It consisted of human-robot dialogue based on two human learners who are friends and introduced a moderate level of difficulty for achieving mastery experience. Overall, we find several interesting design suggestions:

1. For mastery, dialogue design should provide the learner with the impression they are effective; if the robot reaches an answer too quickly, this reduces feelings of effectiveness. Design that incorporates equal question-asking with problem-solving statements can facilitate mastery.
2. For both vicarious experience and social persuasion, our analysis suggests if learners do not feel adequate rapport, they are less likely to have genuine social experiences and this will influence their overall self-efficacy. This implies initially focusing design on fostering rapport.
3. To foster rapport, designing dialogue based on that of friends may produce stronger responses. We are not suggesting designing a robot to act like a long-time friend from the first interaction but targeting initial design strategies to incorporate 'friend-like' moves in initial interactions.
4. Different dialogue designs interact with learners' social predispositions and attitudes towards robots. Problem-solving statements, positivity, and high specificity may increase positive effects for individuals who are less inclined to social interaction with robots and may help influence a positive change in self-efficacy for these individual learners.
5. For individuals with initially low self-efficacy, design for fostering social experiences is more critical. Individuals with initially high self-efficacy responded positively across all phases while individuals with initially low self-efficacy responded positively only to the third design phase.
6. Gesture design should potentially differ depending on the learner's initial level of self-efficacy. We did not perform a full analysis of the design of gesture, keeping emblematic gestures and autonomous life consistent. However, individuals with lower self-efficacy strongly disliked Nico's autonomous life movement, while individuals with high self-efficacy preferred it.

Moving beyond the specific design insights presented in this paper, we see Nico as a platform for exploring the larger space of design questions related to the effects of dialogue and gesture on social learning experiences. This platform allows us to explore small variations in the design of dialogue and ultimately the combination of dialogue and gesture that could have a large impact on learning and motivation, through the creation of social experiences which can help learners develop self-efficacy.

References

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*(3), 586-598.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psych. Review*, *84*(2), 191.

Cassell, J., Gill, A. J., & Tepper, P. A. (2007). Coordination in conversation and rapport. *Proceedings of the Workshop on Embodied Language Processing – Embodied NLP 07*.

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 & Technology, 18*(4), 334-352.

Chou, C., Chan, T., & Lin, C. (2003). Redefining the learning companion: the past, present, and future of educational agents. *Computers & Education, 40*(3), 255-269.

Frager, S., & Stern, C. (1970). Learning by Teaching. *The Reading Teacher, 23*(5), 403-417.

Friday Inst. for Edu. Innovation. 2008. Student Attitudes toward STEM Upper Elementary School Students.

Graesser, A., & Person, N. (1994). Question Asking during Tutoring. *AERJ, 31*(1), 104–137.

Griffin, M. M., & Griffin, B. W. (1998). An Investigation of the Effects of Reciprocal Peer Tutoring on Achievement, Self-Efficacy, & Test Anxiety. *Contemporary Educational Psychology, 23*(3), 298-311.

Gulz, A., Haake, M., & Silvervarg, A. (2011). Extending a Teachable Agent with a Social Conversation Module – Effects on Student Experiences & Learning. *Lecture Notes in CS AI in Education,* 106-114

Hake, R. R. (2002). Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pre-test scores on… *Physics Education Research Conference,* 1-14.

Jacq, A., Lemaignan, S., Garcia, F., Dillenbourg, P., & Paiva, A. (2016). Building successful long child-robot interactions in a learning context. *2016 11th ACM/IEEE HRI*. IEEE Press, 239–246.

Kauchak, D. P., & Eggen, P. D. (1993). *Learning and teaching.* New York: Allyn Bacon.

Leelawong, K., & Biswas, G. (2008). Designing Learning by Teaching Agents: The Betty's Brain System. *International Journal of AI in Education,* 18(3), 181-208.

Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A., & Dillenbourg, P. (2016). Learning by Teaching a Robot: The Case of Handwriting. *IEEE Robotics & Automation Magazine, 23*(2), 56-66.

Levine, M. D. (2002). *Educational care: a system for understanding and helping children with learning differences at home and in school*. Cambridge, MA: Educators Pub. Service.

Liu, P., Glas, D. F., Kanda, T., Ishiguro, H., & Hagita, N. (2013). It's not polite to point: Generating socially-appropriate deictic behaviors towards people. *2013 8th ACM/IEEE HRI.*

Lubold, N., Walker, E., & Pon-Barry, H. (2016). Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. *2016 11th ACM/IEEE HRI*. IEEE Press, 255–262.

Madaio, M. A., Ogan, A., & Cassell, J. (2016). The Effect of Friendship and Tutoring Roles on Reciprocal Peer Tutoring Strategies. *Intelligent Tutoring Systems Lecture Notes in Computer Science,* 423-429.

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative data analysis*. Sage.

Ogan, A., Finkelstein, S., Mayfield, E., Dadamo, C., Matsuda, N., & Cassell, J. (2012a). "Oh, dear Stacy!" Social interaction, elaboration, and learning with teachable agents. *Proc. of the 2012 ACM CHI '12.*

Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012b). Rudeness and Rapport: Insults and Learning Gains in Peer Tutoring. *Intelligent Tutoring Systems Lecture Notes in CS,* 11-21.

Pampaka, M., Williams, J., & Hutcheson, G. (2012). Measuring students' transition into university and its association with learning outcomes. *British Education Research Journal, 38(6),* 1041-1071.

Pareto, L., Arvemo, T., Dahl, Y., Haake, M., & Gulz, A. (2011). A Teachable-Agent Arithmetic Game's Effects on Mathematics Understanding, Attitude and Self-efficacy. *Lecture Notes in CS AI in Edu.,* 247-255.

Robinson, D. R., Schofield, J. W., & Steers-Wentzel, K. L. (2005). Peer and Cross-Age Tutoring in Math: Outcomes and Their Design Implications. *Educational Psychology Review, 17*(4).

Roscoe, R. D., & Chi, M. T. (2007). Understanding Tutor Learning: Knowledge-Building and Knowledge-Telling in Peer Tutors Explanations and Questions. *Review of Educational Research, 77*(4), 534-574.

Tanaka, F., & Matsuzoe, S. (2012). Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning. *Journal of Human-Robot Interaction,* 78-95.

Tickle-Degnen, L., & Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry, 1*(4), 285-293.

Walker, E., Girotto, V., Kim, Y., & Muldner, K. (2016). The Effects of Physical form and Embodied Action in a Teachable Robot for Geometry Learning. *2016 IEEE 16th - ICALT*. IEEE, 381–385.

Wallace, R. S. (2003). *The Elements of AIML Style*. ALICE A.I. Foundation, Inc.

