PROJECT REPORT

Affect detection in math tutoring

Nele Albers, Joanna Mania, Maria Tsfasman

1 Introduction

The tutoring of the twenty-first century has been slowly transforming from a transitional lecture classrooms filled with students, to a project-based personalized learning process led by artificial agents (Kokotsaki et al., 2016). The traditional teaching methods have been challenged in the past decade due to continuously improving capabilities of social robots (Mubin et al., 2013; Brown and Howard, 2014). They have proved to convey learning benefits (Wainer et al., 2006; Leyzberg et al., 2012), elicit emotional expressions (Spaulding et al., 2016), evoke curiosity and growth mindset (Park et al., 2017; Gordon et al., 2015). They might take three roles: peers (Okita et al., 2009), tutor interacting with students, in one-on-one setup (Short et al., 2014) as well as in the front of the classroom (Sisman et al., 2019), or educational assistants that provide answers to straightforward and common questions (Rosenberg-Kima et al., 2019). Science, languages, nutrition education and math are just a few examples of their applications in the educational setting (Brown and Howard, 2014; Kory and Breazeal, 2014; Shiomi et al., 2015).

This report presents a development of a socially aware conversational agent taking a role of a math tutor that accommodates its behaviour to the performance and mood of the student. Taking into account the sensitive nature of the application, we strove to create a highly explainable and predictable system that allows high level of control over the interaction to prevent from undesired and hurtful actions towards the users.

The report is structured as follows. Firstly, we present a motivation for the present study. Next section covers the implementation, where key functions in the interaction design, emotional state and empathetic feedback are described. Next, the evaluation of the affect detection and social presence of the agent are presented. Further, we provide a discussion of the results, limitations and suggestions for the future study, to end with concluding remarks.

2 Motivation

Rendering dialog systems socially aware is important both for a conversation itself and the (long-term) impact of the conversation on the user. Regarding the former, (van Turnhout et al., 2008) point out that a conversational agent needs to be aware of the communicative context to understand who is being addressed. Otherwise, the agent may incorrectly react to utterances that are not directed at it, which in turn might disrupt human-human conversations. Furthermore, (van Turnhout et al., 2008) also stress that a system needs to display active listening behavior such as via nodding. The reason is that the human conversation partner may otherwise think that the system is inattentive and stop conversing.

Besides the conversation itself, the (long-term) effect of a conversation also is impacted by the social awareness of a dialog system. For example, prior work showed the positive impact of rapport on Math performance (Cassell et al., 2007), learning gains (Sinha & Cassell, 2015), supporting the user's goals (Papangelis et al., 2014) and negotiation (Drolet & Morris, 2000). Similar results hold for convergence (Sinha & Cassell, 2015; Friedberg et al., 2012), which measures the degree to which dialog partners become more similar to each other during a dialog. Notably, both rapport and convergence can only be achieved by a dialog system if it is socially aware. For instance, mutual attentiveness and face management are important factors to create rapport (Zhao et al., 2016), both of which depend on social awareness.

In the education context in which this work is placed, feedback and motivation can more specifically be made more effective if a dialog system is socially aware. For example, (Rowe, 2017) shows that positive emotions can lead to better academic learning, whereas negative emotions can have a negative impact on motivation and performance. Thus, if a dialog system is aware of the emotions of its dialog partner, it can deduce how motivated he or she is and react accordingly. Furthermore, since emotions such as joy, anger and embarrassment strongly impact how students react to feedback (Rowe, 2017), feedback can also be made more effective by considering emotions. Besides emotions, however, other factors such as a person's cultural background and whether or not a student's result matches her expectations also need to be considered when giving feedback (Ryan & Henderson, 2018a).

3 Implementation

The implementation started with a schematic dialogue flow (Figure 1) of a desired interaction between the math tutor and a user.

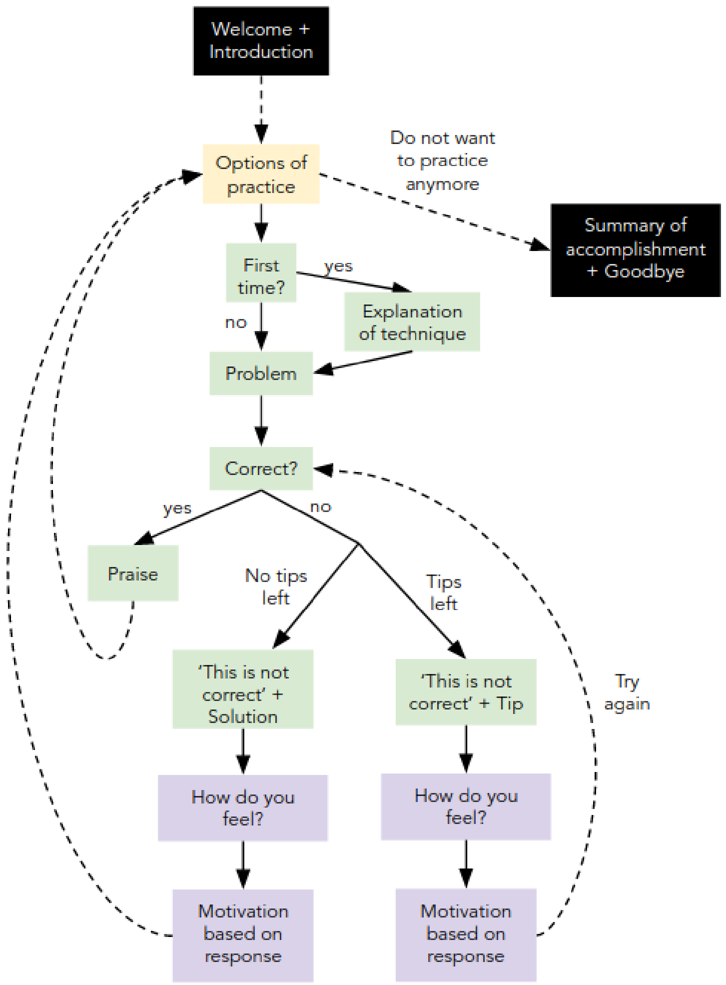

Figure 1: The prototype dialogue flow, based on which the interaction was designed (modifications were made with more details and some changes depending on the implementation effectiveness).

3.1 Key functions in the interaction design

Since the final dialogue design had too many details to show in a readable flow chart, the interaction between main function is show in Figure 2. The main functions included 'emotional state', 'choose exercise', 'do exercise', 'no problem left' (2). The program starts with user's emotion check (described in more detail below). Then, the user proceeds with choosing the exercise, if there are still exercises left the user may try completing the exercise. Throughout the exercise completion the tutor is acting supportive and introduces different tips when the user's answers are incorrect. If there were several incorrect answers the robot proceeds to *emotion check* to find out if the user is still motivated to do more.

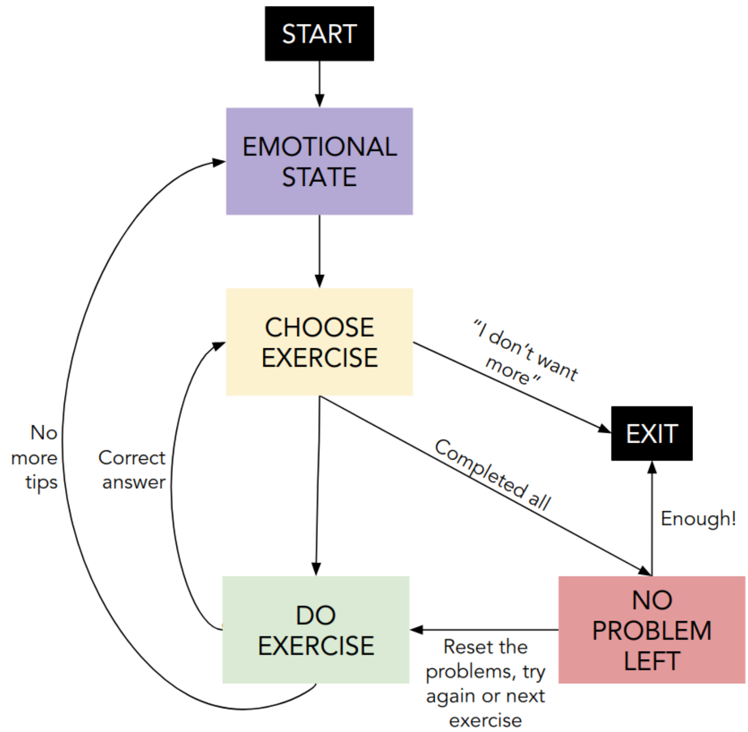

Figure 2: The interaction of resulting key functions in the interaction (the full flow charts of each block can be found here)

3.2 Emotional state and empathetic feedback

Of particular importance for the implemented interactive tutor is its empathy module. The emotional checks are conducted by the means of an explicit question (e.g. "How are you feeling?"), and a sentiment analysis (Figure 3).

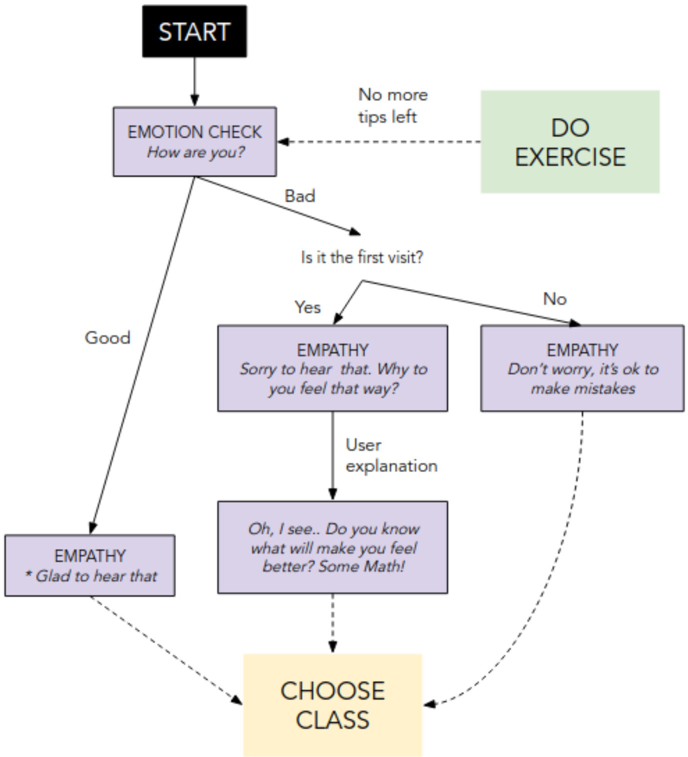

Figure 3: A more detailed scheme of emotional state class

The empathetic feedback includes both verbal and nonverbal signals of encouragement and praise. To achieve maximum explainability and predictability of the system, the algorithm for *emotion check* was sentiment analysis using opinion lexicon keywords. The sentiment lexicon we used consisted of more than 6800 English positive and negative opinion words (Liu et al., 2005). The *emotion check* function parses through student's responses and identifies the number of negative vs positive keywords. If the number of negative words is higher than positive, the empathetic response is triggered - the robot acts sad and produces motivational phrases to cheer the student up (see examples of phrases in Figure 3). Such system adaptation is expected to make the student feel more motivated throughout the lesson.

As a process parallel to verbal and non-verbal expression linked to specific states, a set of auto expressions was implemented to run throughout the whole interaction. Micro gaze while speaking, sporadic looking around, head movements when attending the user or looking away while listening to the user were implemented to make the agent behaviour more alive and natural.

3.2.1 Literature-based empathetic feedback: verbal and non-verbal

In order to maintain student's motivation throughout the lesson, we based the non-verbal and verbal feedback on the educational studies. According to (Wang & Loewen,

2016) there are several common nonverbal signals in supportive human teachers:

(1) Nodding is a common head movement: either nodding repeatedly to confirm a student hypothesis while student says it, or nodding a single time to give emphasis to a word pronunciation made by teacher

(2) Head shaking is a common head movement: analysis shows that it is almost unanimously a way for teachers to disconfirm a student utterance

(3) Most typical affect display involved facial expressions, e.g. eyes wide open when not understanding a student

All those non-verbal behaviours were implemented as feedback to the student during the completion of an exercise. According to (Kim et al., 2007), when the agent "empathetically" matched the learner's emotion, the students showed more interest and higher self-efficacy. In our implementation, therefore the agent acted sad when the user would mention more negative words and happy - when more positive words were used. According to (Babad, 2009) teachers can increase their teaching effectiveness by improving their enthusiasm, immediacy, and expressive style. Therefore, our tutor would smile and act enthusiastic during the lesson except for when the user emotional state would signify the user in a bad mood.

Regarding verbal feedback, Based on (Blanson Henkemans et al., 2009) a tutor agent should:

(1) Express empathy

(2) Cheer and compliment the student when reaching the goals

(3) Give constructive feedback

(4) Be optimistic

Those are similar to the goals of the non-verbal feedback from (Wang & Loewen, 2016). All those are integrated into our tutor. The type of verbal feedback also matters when responding to different emotions of the student. Overall, all research points into direction of empathy: tutor's feedback should match the user's state of mind, express understanding and be overall optimistic(Rowe, 2017; Rowe et al., 2014; Ryan & Henderson, 2018b).

4 Results

4.1 Affect Detection

The detection of the dialog partner's affect is crucial for a tutoring agent to make its feedback and motivational messages accurate and effective. To assess the quality of our agent's affect detection, we presented the agent with 31 affect statements and measured whether or not it correctly identified the affect as positive or negative. Table 1 shows that the agent misclassified merely five of the 31 statements, leading to an overall accuracy of 0.84. Four of these incorrectly classified statements are the only ones that contain the word "not," which reverses the polarity of the affect expressed by the other words in the statement. Hence, our agent is not able to handle such polarity reversal (visible in the attached video). The other misclassified statement is "I'm okay." Arguably, this statement truly expresses neither a positive nor a negative affect and would hence best be classified as neutral. The misclassification can thus be explained by the fact that neutral statements are classified as "negative" by our agent.

Table 1: Affect detection based on 31 samples. True and predicted labels as well as the overall accuracy are shown.

Sentence	True Label	Predicted Label	Correct
Good	Positive	Positive	✓
I'm good	Positive	Positive	✓
Great thanks	Positive	Positive	✓
Not good at all	Negative	Positive	×
Today is not the best day	Negative	Positive	×
Fine thanks	Positive	Positive	✓
Not the best	Negative	Positive	×
I'm very sad today	Negative	Negative	✓
I feel very unmotivated today	Negative	Negative	✓
I'm exhausted	Negative	Negative	✓
Frustrated	Negative	Negative	✓
Stupid	Negative	Negative	✓
Silly	Negative	Negative	✓
I'm stupid	Negative	Negative	✓
I'm okay	Positive	Negative	×
Sad	Negative	Negative	✓
Excited	Positive	Positive	✓
I'm excited	Positive	Positive	✓
Fantastic	Positive	Positive	✓
Okay	Negative	Negative	✓
I'm fine	Positive	Positive	✓
I feel amazing	Positive	Positive	✓
It is not going good	Negative	Positive	×
It's going pretty good	Positive	Positive	✓
I am feeling good	Positive	Positive	✓
Today I'm exhausted	Negative	Negative	✓
I feel really stressed out	Negative	Negative	✓
Very sad	Negative	Negative	✓
I am dissatisfied	Negative	Negative	✓
I feel awesome	Positive	Positive	✓
Today I feel happy	Positive	Positive	✓
		Accuracy	**0.84**

4.2 Social Presence

Affect detection is solely one aspect that influences the social awareness of a conversational agent. To determine the overall social awareness of our agent as perceived by its conversation partners, we employ the social presence questionnaire by (Harms & Biocca, 2004). This questionnaire captures the "degree of initial awareness, allocated attention, the capacity for both content and affective comprehension, and the capacity for both affective and behavioral interdependence" (Harms & Biocca, 2004) with the conversational agent, all of which are important facets of a successful tutoring agent. This questionnaire was given to three participants after conversing with the agent for about three minutes. While such a small sample size does not allow for a reliable measure of the six constructs, it does give an indication of the strengths and weaknesses of our agent.

Our results suggest that while our conversational agent achieves relatively high degrees of initial awareness (M: 5.56, SD: 1.71) and perceived behavioral interdependence (M: 5.22, SD: 1.27), its scores for perceived emotional interdependence (M: 2.94, SD: 1.96) and perceived affective understanding (M: 3.83, SD: 1.80) are rather low. This indicates that it is especially the agent's way of dealing with affect that can be improved upon. The scores for perceived message understanding (M: 4.44, SD: 1.80) and attentional allocation (M: 4.61, SD: 2.41) fall in between. Given that there is no general threshold for these constructs from which onward a conversational agent is considered "good," it is, however, not clear whether the ratings obtained for the different constructs are directly comparable in this fashion. For example, it is not certain that a score of 5.56 for co-presence actually is better than a score of 4.61 for attentional allocation.

In addition, a reliable evaluation of our results is further hindered by the relatively large differences between the ratings given by our three participants. Across the entire questionnaire, we observe a Cronbach's α of 0.46, which indicates only moderate agreement according to Cohen (McHugh, 2012). Looking at the six constructs separately, we further find that there is no agreement for the constructs perceived affective understanding, perceived emotional interdependence and perceived behavioral interdependence, moderate agreement for attentional allocation and co-presence, and substantial agreement for perceived message understanding. This suggests that there are strong differences between individuals with regards to affective comprehension and affective and behavioral interdependence. Furthermore, only for perceived message understanding do we find strong agreement between our participants.

5 Discussion

Subsequently, we discuss both the strengths and the weaknesses of our system.

5.1 Strengths

An evident strength of our system is its relatively high co-presence score. This means that its verbal and nonverbal behavior allow the agent to create a mutual initial awareness. As previously noted, however, it is not evident whether the scores obtained for the six constructs of the social presence questionnaire are directly comparable. It would hence be very useful if there were general guidelines as to which scores for the different constructs constitute a "good" level of social presence.

A second strength of our conversational agent is its high affect detection accuracy of 0.84. While our affect detection is rather simple in that it distinguishes solely two broad types of affect, this also reduces the chance of misclassification. Even though the responses given by our agent are hence relatively vague, they are at least not entirely inappropriate. The latter is very important to make the conversational partner feel understood and to not hurt or upset him or her.

5.2 Limitations

Our system has several limitations that are important to point out. One important one is that we observe rather low mean ratings for the way our conversational agent

Table 2: Questionnaire items from the social presence questionnaire (Harms & Biocca, 2004) and the ratings given by three participants on 7-point Likert scales. The ratings were given after talking to the agent for about 3 minutes.

Item	Rating 1	Rating 2	Rating 3	
Co-presence				**Mean: 5.56, SD: 1.71, Cronbach's α: 0.51**
1	7	4	7	
2	7	5	3	
3	7	7	3	
4	7	7	3	
5	7	7	7	
6	5	4	3	
Attentional Allocation				**Mean: 4.61, SD: 2.41, Cronbach's α: 0.43**
7	1	6	4	
8	3	7	6	
9	7	7	2	
10	7	7	4	
11	1	2	7	
12	2	1	7	
Perceived Message Understanding				**Mean: 4.44, SD: 1.80, Cronbach's α: 0.76**
13	4	1	2	
14	7	4	7	
15	7	4	7	
16	4	5	3	
17	1	5	4	
18	4	5	4	
Perceived Affective Understanding				**Mean: 3.83, SD: 1.80, Cronbach's α: -0.14**
19	7	5	4	
20	5	2	2	
21	4	4	2	
22	5	7	2	
23	4	6	2	
24	5	1	2	
Perceived Emotional Interdependence				**Mean: 2.94, SD: 1.96, Cronbach's α: -0.32**
25	4	1	1	
26	4	1	5	
27	6	1	1	
28	5	1	2	
29	6	6	2	
30	2	1	4	
Perceived Behavioral Interdependence				**Mean: 5.22, SD: 1.27, Cronbach's α: -0.59**
31	6	5	7	
32	6	2	7	
33	5	6	5	
34	4	5	5	
35	5	7	5	
36	6	3	5	

handles affect. This likely is due to the fact that the agent considers affect only at specific points during the interaction. That is, the agent asks its conversation partner how he or she is feeling solely at the start of the conversation and after exercises were attempted unsuccessfully. We chose to consider affect at these points during the interaction, because the literature suggests that those are important situations to consider in a tutoring interaction (Rowe, 2017; Ryan & Henderson, 2018a). Our experiments, however, show that especially negative affect such as frustration can also arise during other parts of the conversation. One participant, for example, was frustrated that the question was read out too quickly and that an insufficient amount of explanation was given beforehand. By not reacting to this frustration until after the participant had given a wrong answer, the agent likely made the participant feel not understood emotionally.

Moreover, since our agent distinguishes only two types of affect, some of its reactions could be improved upon. If, for example, a person says ”I'm okay,” the agent may reply ”I'm happy to hear that.” Arguably, a more appropriate response given the expressed neutral affect would be along the lines of ”Ah, that's not too bad.” Distinguishing more levels of affect besides ”positive” and ”negative” would hence be useful. Yet, it is crucial that increasing the number of affect categories does not lead to higher misclassification rates. Clearly, giving a vague response such as ”I'm sorry to hear that” to any negative affect is better than replying with a more specific but wrong response, such as ”I'm sorry to hear that you are angry” if the user is in fact sad rather than angry, for example.

Another important observation we made is that the given questions were rather difficult and required more time to solve than was given by the agent. Hence, especially participants with lower Math skills were not able to solve the questions and got frustrated. If this system was to be used in practice, exercises of different difficulty levels should thus be offered with the option to either let the participant choose a level of difficulty or make an informed decision about an appropriate level of difficulty for a given person. This would likely also reduce the problem of unaccounted negative affect at the time when a question is posed.

Lastly, a reliable evaluation of our system is hindered by the strong differences between the ratings given by the three participants, especially for perceived affective and emotional interdependence and perceived affective understanding. These differences could be due to varying cultural backgrounds, since our participants were from the culturally diverse countries Nepal, Poland and Russia, as well as different genders. For perceived emotional interdependence, for example, our only male and Asian participant gave much higher ratings on average (M: 4.5, SD: 1.38) than our two female and European participants (M: 1.83 and 2.5, SD: 1.86 and 1.5). A similar observation holds for perceived affective understanding, where we obtained a mean rating of 4.67 by our male participant and mean ratings of 3.17 and 3.67 by our two female participants. A systematic study with more participants would be needed to determine whether these observed differences are solely due to chance or indeed caused by differences between genders, cultural backgrounds or other factors such as Math skills, familiarity with English or previous experience with conversational agents.

6 Conclusion

To summarise, we developed a mathtutor system able to track student's emotion through keyword extraction. The mathtutor's verbal and non-verbal feedback to the student is firmly based on educational research to achieve better results with keeping a student motivated and avoid user frustration with the subject. The system strives for maximum explainability and predictability over generated behaviours and therefore is mostly rule-based. Although the system itself does not use any innovative approaches to affect detection and dialog management, the results seem rather promising. Our short user study showed good ratings of social presence, perceived behavioural interdependence and attentional allocation. Our affect detection showed relatively high accuracy rates although being very simple. Since the affect was only detected in specific moments of the lesson, the users ratings of affect management did not show to be very high. This might be improved by measuring user's affect throughout the whole interaction.

There are some limitations, such as affect detection throughout the whole session rather than only before and after a task could work better. It is important to mention that the system needs further testing with a larger number of participants to make any definitive conclusions.

Our results might be interpreted in that even a rule-based system with a simple keyword sentiment analysis might be sufficient for a virtual tutor, if its verbal and nonverbal behaviour is firmly based on education research. We suggest that although innovative approaches seem to give more accurate results, it does not always mean that the system would be perceived as more sophisticated by users. Yet, the loss of explainability and predictability is something to take into account when using a machine learning model for affect detection or automatically generated verbal responses.

7 References

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving schools*, *19*(3), 267–277.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J.-J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, *1*(209-0015), 13.

Brown, L. N., & Howard, A. M. The positive effects of verbal encouragement in mathematics education using a social robot. In: *2014 ieee integrated stem education conference*. IEEE. 2014, 1–5.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. The role of physical embodiment in human-robot interaction. In: *Roman 2006-the 15th ieee international symposium on robot and human interactive communication*. IEEE. 2006, 117–122.

Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. The physical presence of a robot tutor increases cognitive learning gains. In: *Proceedings of the annual meeting of the cognitive science society. 34*. (34). 2012.

Spaulding, S., Gordon, G., & Breazeal, C. Affect-aware student models for robot tutors. In: *Proceedings of the 2016 international conference on autonomous agents & multiagent systems*. 2016, 864–872.

Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., & Breazeal, C. Growing growth mindset with a social robot peer. In: *Proceedings of the 2017 acm/ieee international conference on human-robot interaction*. 2017, 137–145.

Gordon, G., Breazeal, C., & Engel, S. Can children catch curiosity from a social robot? In: *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*. 2015, 91–98.

Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. Learning together: Asimo developing an interactive learning partnership with children. In: *Ro-man 2009-the 18th ieee international symposium on robot and human interactive communication*. IEEE. 2009, 1125–1130.

Short, E., Swift-Spong, K., Greczek, J., Ramachandran, A., Litoiu, A., Grigore, E. C., Feil-Seifer, D., Shuster, S., Lee, J. J., Huang, S., et al. How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play. In: *The 23rd ieee international symposium on robot and human interactive communication*. IEEE. 2014, 924–929.

Sisman, B., Gunay, D., & Kucuk, S. (2019). Development and validation of an educational robot attitude scale (eras) for secondary school students. *Interactive Learning Environments*, *27*(3), 377–388.

Rosenberg-Kima, R., Koren, Y., Yachini, M., & Gordon, G. Human-robot-collaboration (hrc): Social robots as teaching assistants for training activities in small groups. In: *2019 14th acm/ieee international conference on human-robot interaction (hri)*. IEEE. 2019, 522–523.

Kory, J., & Breazeal, C. Storytelling with robots: Learning companions for preschool children's language development. In: *The 23rd ieee international symposium on robot and human interactive communication*. IEEE. 2014, 643–648.

Shiomi, M., Kanda, T., Howley, I., Hayashi, K., & Hagita, N. (2015). Can a social robot stimulate science curiosity in classrooms? *International Journal of Social Robotics*, *7*(5), 641–652.

van Turnhout, K., Terken, J., & Eggen, B. Designing socially aware conversational agents (E. André, L. Dybkjær, W. Minker, H. Neumann, R. Pieraccini, & M. Weber, Eds.). In: In *Perception in multimodal dialogue systems* (E. André, L. Dybkjær, W. Minker, H. Neumann, R. Pieraccini, & M. Weber, Eds.). Ed. by André, E., Dybkjær, L., Minker, W., Neumann, H., Pieraccini, R., & Weber, M. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, 256–267.

Cassell, J., Gill, A., & Tepper, P. Coordination in conversation and rapport. In: *Proceedings of the workshop on embodied language processing*. 2007, 41–50.

Sinha, T., & Cassell, J. Fine-grained analyses of interpersonal processes and their effect on learning. In: *International conference on artificial intelligence in education*. Springer. 2015, 781–785.

Papangelis, A., Zhao, R., & Cassell, J. Towards a computational architecture of dyadic rapport management for virtual agents. In: *International conference on intelligent virtual agents*. Springer. 2014, 320–324.

Drolet, A. L., & Morris, M. W. (2000). Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Journal of Experimental Social Psychology*, *36*(1), 26–50.

Friedberg, H., Litman, D., & Paletz, S. B. Lexical entrainment and success in student engineering groups. In: *2012 ieee spoken language technology workshop (slt)*. IEEE. 2012, 404–409.

Zhao, R., Sinha, T., Black, A. W., & Cassell, J. Automatic recognition of conversational strategies in the service of a socially-aware dialog system. In: *Proceedings of the*

17th annual meeting of the special interest group on discourse and dialogue. 2016, 381–392.

Rowe, A. D. Feelings about feedback: The role of emotions in assessment for learning. In: *Scaling up assessment for learning in higher education*. Springer, 2017, pp. 159–172.

Ryan, T., & Henderson, M. (2018a). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education*, *43*(6), 880–892.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web.

Wang, W., & Loewen, S. (2016). Nonverbal behavior and corrective feedback in nine esl university-level classrooms. *Language Teaching Research*, *20*(4), 459–478. https://doi.org/10.1177/1362168815577239

Kim, Y., Baylor, A., & Shen, E. (2007). Pedagogical agents as learning companions: The impact of agent emotion and gender. *Journal of Computer Assisted Learning*, *23*, 220 –234. https://doi.org/10.1111/j.1365-2729.2006.00210.x

Babad, E. (2009). Teaching and nonverbal behavior in the classroom. In L. J. Saha & A. G. Dworkin (Eds.), *International handbook of research on teachers and teaching* (pp. 817–827). Springer US. https://doi.org/10.1007/978-0-387-73317-3_52

Blanson Henkemans, O., Boog, P., Lindenberg, J., Mast, C., Neerincx, M., & Schonk, J. (2009). An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. *Technology and health care : official journal of the European Society for Engineering and Medicine*, *17*, 253–67. https://doi.org/10.3233/THC-2009-0545

Rowe, A. D., Fitness, J., & Wood, L. N. (2014). The role and functionality of emotions in feedback at university: A qualitative study. *The Australian Educational Researcher*, *41*(3), 283–309. https://doi.org/10.1007/s13384-013-0135-7

Ryan, T., & Henderson, M. (2018b). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education*, *43*(6), 880–892. https://doi.org/10.1080/02602938.2017.1416456

Harms, C., & Biocca, F. (2004). Internal consistency and reliability of the networked minds measure of social presence.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276–282.