Self-Assessment Through Interactive In-Action Reflections to Improve Interpersonal Skills Training

Diego Rivera-Gutierrez, Benjamin Lok
Department of CISE
University of Florida
Gainesville, FL, USA
Email: {djrg, lok}@cise.ufl.edu

Andrea Kleinsmith
Department of Information Systems
University of Maryland, BC
Baltimore, MD, USA
Email: andreak@umbc.edu

Gail Childs, Roberta Pileggi
College of Dentistry
University of Florida
Gainesville, FL, USA
Email: {gchilds, rpileggi}@dental.ufl.edu

Abstract—People often under or over-estimate their performance on interactive learning experiences with virtual agents, especially when it comes to interpersonal skills such as empathy. To generate more accurate self-assessment of performance, we propose the use of in-action reflective learning opportunities during interactive learning experiences with virtual agents. We conducted a user study in which third-year dental students (n=58) participated in an interactive learning experience that required them to demonstrate empathy towards a virtual agent playing the role of a patient. During the interaction, an in-action reflective learning intervention prompted the students to self-assess their performance with regards to empathy. Our results show that students’ self-assessment correlates to an assessment performed by outside raters, and that students were significantly more empathetic on a second opportunity to demonstrate empathy to the virtual patient after having performed their reflection.

Keywords—Virtual patients; self-assessment; reflective learning; in-action; empathy; virtual tutors

I. INTRODUCTION

Interpersonal skills such as being able to communicate empathy to a patient can be critical to the success of a health sciences professional [7]. Because interpersonal skills can be so critical, interactive learning experiences with virtual agents playing the role of patients have been used to give learners a safe environment to practice them [5], [6]. Continued practice with such virtual patients can help learners be better prepared for future experiences with real patients. Additionally, recognizing their shortcomings during those interactions can guide learners to better outcomes. However, recognizing your own shortcomings is of critical importance in any learning activity. The work presented in this paper looks at a novel way in which a virtual instructor can be used in virtual agent experiences to help students recognize their limitations, while at the same time guiding them in how to improve their skills.

Many approaches of personal change involve a first step of recognizing or identifying the existence of a problem. In order to recognize those problems, students need to become critical in assessing their own performance. The activity of having students grade or assess their own performance is commonly referred to as self-assessment [1]. However, research shows that when self-assessing their performance, students often under or over-estimate how well they performed [1]. Specifically, self-assessing performance on interpersonal skills, such as demonstrating empathy towards another person or a virtual agent, can be very challenging.

Interactive learning environments with virtual agents have been used to train these interpersonal skills in the past in fields such as military field operations [6], and health sciences [8]. In these sort of training scenarios, self-assessment has been mostly used in the context of after-action reviews [6], [8]. In these after-action reviews, the learners are asked to review their own performance after the interaction with the virtual agent is over. While such approaches can lead students to important self realizations [8], they require the presentification of a past experience which relies on the reliability of the student’s memory [10] (i.e. the students need to recall or be reminded of the whole context of what was happening at a critical moment of the learning exercise). Important cognitive information related to the context of the situation and their cognitive state can be lost in this process of presentification.

We propose the use of in-action reflections led by a virtual tutor as an instructional scaffold to promote self-assessment and to help the learner identify their own shortcomings. These in-action reflections pause the interactive learning experience right after the critical moment of interest. This pause is designed with two purposes: (1) to present the student with questions that guide them through an introspection process of self-assessment, and (2) to provide the student with key elements that can make their future responses better. After this learning intervention, the training session continues. We used this approach in a user study in which 58 dental students self-assessed their communication of empathy to a virtual patient using a modified version of the Empathic Communication Coding Scale (ECCS) [2]. The main findings of this study are: (1) The students’ self-assessment shows a moderate correlation ($r = 0.48$) with third-party ratings of their empathy. The correlation shows an indication that the students were relatively accurate in their self-assessment of how empathetic they were to the virtual patient. (2) The students were significantly more empathetic to the patient after their in-action self-assessment of their empathy ($p < 0.01$).

II. INTERPERSONAL SKILLS TRAINING WITH VIRTUAL AGENTS

Interactive learning environments with virtual agents have become an educational tool for students to practice interpersonal skills including military negotiation [3], [6] and medical interviewing and diagnosis [8]. In these scenarios, virtual agents present learners with a low-pressure, low-risk...
environment to practice their interpersonal skills while limiting the consequences of potential mistakes during training [9]. While these virtual agent experiences by themselves represent an important educational tool, educational research suggests that a critical component to improving in these skills is the metacognitive process of reviewing those experiences [3]. This review process allows learners to identify deficiencies in their performance and specific skills that they need to improve on.

Implicitly, most virtual agent experiences engage learners in this metacognitive process of reviewing their performance. However, explicitly engaging students in metacognitive processes is not as common. In virtual agent literature the most common way of explicitly doing so has been through After-Action Reviews (AAR). An AAR is an exercise in which the learner revisits event after this has occurred. This exercise allows the user to better understand the exercise, helping them achieve the required training objectives. AARs have been used to provide insight of how the behavioral model of the virtual agents works [3] and to revisit an interaction from the point of view of the virtual agent [8].

More recently, Rivera-Gutierrez et al. [9] proposed a reflective practicum framework for virtual agent experiences that encourages learners to engage in a metacognitive process of reflection. This framework introduces a virtual tutor that guides students through the reflective process at key points of the interaction. The framework features the opportunity to generate in-action reflections. In-action reflections happen during the interaction with a virtual agent, pausing the interaction to give the learner an opportunity for introspection. In-action reflections thus eliminate the need for the student to recall the events from memory by bringing the educational intervention to the key moment in the interaction. This paper leverages this reflective practicum framework [9] to generate a key moment of in-action reflection that guides learners in self-assessing their performance regarding the empathy they communicated to a virtual patient. We study the accuracy of this self-assessment as well as potential behavior changes in future opportunities to demonstrate empathy towards the virtual patient.

III. SELF-ASSESSMENT

Self-assessment is the activity of having students grade or assess their own performance [1]. Self-assessment is therefore a metacognitive activity in which students try to understand their own cognitive process during a learning experience and quantify their performance in some metric. The ability to accurately self-assess one’s performance can be a critical part of reaching learning objectives in a task. However, measuring the level of accuracy of a student’s self-assessment is not an easy task. Research in self-assessment accuracy usually follows a correlational model. This correlational model is a single condition empirical design in which the student’s self-assessment is compared to a “gold standard” measure [4]. This gold standard is usually an assessment from an expert or third party raters. The gold standard metric is later correlated to the student’s self-assessment to measure the accuracy of the student’s assessment. High correlations are taken as good measures of self-assessment.

To address criticism to this paradigm, Gruppen and Regehr [4] have outlined a number of suggestions to deal with some of this potential pitfalls. Their suggestions include: (1) Providing students with specific anchors (i.e. examples and definitions) through the self-assessment process to provide a more stable use of the scale. (2) The use of multiple expert raters to improve the reliability of the standard.

Research suggest that engaging students in this process of self-assessment is an important tool that can help them develop their interpersonal skills [6]. However, when training these interpersonal skills through virtual agent experiences, the process of self-assessment has rarely been explicit. Even when made explicit, through after-action reviews, self-assessment has not been quantified and its accuracy has not been studied. We suggest the use of in-action reflections for self-assessment of interpersonal skills during virtual agent experiences. In this kind of self-assessment a virtual tutor asks a sequence of close-ended questions that guide the process of self-assessment. These close-ended questions have to be simple enough for the student to self-assess without having deep knowledge of the metric being used to measure the particular interpersonal skill. To allow this simplicity, the questions should include anchors as suggested by Gruppen and Regehr [4]. The prompts have to define the concepts the learner might not be familiar with and provide examples to illustrate them. We suggest that if these prompts are simple enough and designed correctly, they not only provide an opportunity for self-assessment, but can also be a roadmap that allows the student to identify the strengths and weaknesses of their performance.

We have applied this paradigm to a virtual agent experience to teach dental school students about communicating empathy. Figure 1 shows an example of one of the self-assessment prompts created for this scenario. The prompt 1) reminds the learner what their answer to the patient’s concerns was, 2) presents them with a dimension of empathy that they are expected to self-assess, and 3) provides options for the student to quantify his or her performance.

IV. EMERGENCY TOOTH PAIN VIRTUAL PATIENT

We applied the in-action reflection for self-assessment paradigm to a virtual patient experience targeted to teach
dental students how to interview, diagnose and be empathetic to a patient presenting with extreme tooth pain. The virtual patient in the scenario is called “Maria”. Maria was developed in collaboration with the University of Florida Dental School, and she is a 29 year old Hispanic female suffering from severe tooth pain. Maria is seeking emergency attention from the student. Students interact with Maria through a web-based interface. The interface allows learners to type their questions for Maria using natural language. Maria responds to student questions using pre-recorded speeches and animations. Maria can respond to 2532 questions with 417 unique responses. On average students spend 12 minutes interacting with Maria.

As part of the learning objectives for the experience, learners are expected to communicate empathy towards Maria. Empathy is a complex concept and can be defined in many ways: cognitive, affective or both [5]. In this case, the task that is being assessed is primarily a cognitive task: students are expected to identify Maria’s concerns and find an appropriate response that would be effective at addressing those concerns. Learners are not being assessed or being asked to assess whether or not they “felt” empathy towards Maria.

Maria provides the learner with two opportunities to communicate physician empathy: (1) Payment: Four minutes into the interaction Maria says: “Doctor, I don’t know if I’ll be able to pay for this treatment, are there other options?” (2) Bad experience: Ten minutes into the interaction Maria says: “Doctor, Is this a long procedure? I had a bad dental experience when I was a child and have been fearful of going to the dentist since then.” After Maria expresses each of those concerns, a dialog box allows the learners to respond empathetically to her concerns.

The emotional intensity of these two empathetic opportunities was validated following the process described by Kleinsmith et al. [5]. 20 non-expert raters were recruited online, and instructed to rate the emotional intensity of each opportunity on a five-point bipolar scale. The average measure intraclass correlation coefficient (i.e., consistency of the raters’ values for the opportunities) was 0.710, \( p < 0.001 \) and the median intensity was a 3 for both opportunities. This indicates that both opportunities have similar emotional intensity.

Immediately after responding to Maria’s concerns for the first empathetic opportunity regarding payment, the learners are introduced to an in-action reflective moment. Table I shows the questions that were used during the reflective moment and for the process of self-assessment. The three closed-ended questions in the reflective moment constitute a modified version of the Empathic Communication Coding Scale (ECCS) [2]. The ECCS is a validated seven-point ordinal scale (0 being the least empathetic and 6 being the most empathetic) according to which outside raters can rate a physician’s response to an empathetic statement from a patient. The ECCS scale has been used in the past to measure how empathetic students are towards virtual patients in training scenarios aimed to teach empathy [5]. The scale was modified to allow self-assessment from the students. In order to modify the scale, we identified the three main components described as being part of empathic communication: acknowledgment, pursuit, and confirmation or shared feeling. While the ECCS quantifies all three elements together, each one of them can be quantified individually by a student as being present or absent from their empathetic response. Following the rules of the scale provided in the literature these individual assessments can later be mapped back into the full seven-point scale. This breakdown of the scale allows a learner to more quickly assess their performance, while at the same time providing them with the knowledge of what is expected of a good empathetic response towards a patient.

V. User Study

Using the emergency tooth pain scenario with in-action self-assessment we conducted a user study. Following the correlational model of self-assessment research [4], the user study was designed as an empirical study in which all participants were assigned to a single treatment condition without a control group. For this user study, the self-assessment was limited to the empathy displayed on the first empathetic opportunity presented by Maria regarding her concern about not having health insurance, and not knowing how she will be able to afford treatment.

The user study had two hypotheses. **Primary hypothesis:** The ECCS empathy ratings from outside raters will be strongly correlated to the self-assessment performed by the learners during the in-action reflection. **Secondary hypothesis:** The learners will show an increase in empathy between the first and second empathetic challenges.

A. Metrics

Self-assessed ECCS rating: The answers provided by the students to the closed-ended questions of the in-action self-
assessment of the first empathetic moment were mapped back into the full seven-point ECCS scale.

**ECCS ratings from outside raters:** Three expert raters rated both empathetic opportunities included in the interaction.

### B. Participants

A total of 65 (26 male) third year dental students from the University of Florida Dental School participated in the study as part of their evidence-based endodontics course. Participants received extra credit in the class for participating in the exercise. The students had an average age of 26.3 years with a standard deviation of 2.7 years.

### VI. Analysis and Results

Our analysis of the data confirmed both of our hypotheses. First, the data showed that there was a good correlation ($r = 0.48$) between the ratings from outside raters and the self-assessment performed by the students with regards to the level of empathy the students communicated to Maria during the first opportunity to display empathy. Additionally, while 57% of the participants indicated that they would not change their empathetic answer to Maria during the reflective moment, the data also showed that students were rated as being significantly more empathetic ($p < 0.01$) in their responses to the second empathetic moment than the first by the outside raters.

The students’ empathetic responses were coded using the full 7-point ECCS scale by three expert raters. The raters are researchers with experience in human–virtual patient interactions, but were blind to the purpose of the study and the self-ratings provided by the students in their self-assessment. The expert raters have experience rating empathetic answers and have gone through training in the scale together.

The experts rated the answers provided by all 65 students to both empathetic moments. During this process the raters were blinded to all identifying information of the participants, including the pairings between the first and second empathetic responses. The experts rated the responses individually without knowledge of the other coders’ ratings. The interrater reliability of the coding of the empathetic responses was computed using a single intraclass correlation coefficient (ICC) which included the ratings for both empathetic moments. Very high interrater reliability was achieved, ICC$(3,3) = .842$, ($p < .001$).

This ratings performed by outside raters were taken as the gold standard for the students in the empirical design of the user study. For this particular analysis, high levels of agreement were necessary between the raters. Therefore, in cases of disagreement, the mode value was taken since at least two of the three raters agreed that was the more accurate assessment of the empathetic content in the response. In the seven cases where all three raters disagreed on the rating for a particular response, the data from that particular interaction was excluded from the correlation analysis to maintain the fairness of the gold standard. Therefore, only data from 58 students was used for the correlation analysis. The self-assessment performed by the students was moderately correlated to the mode rating provided by the outside raters, $r(56) = 0.48$, $p < 0.01$. Figure 2 shows a jitter plot of the correlation between these ratings.

Additionally, we performed a paired samples t-test between the ratings provided by the outside raters for the first and second empathetic moments. This analysis showed that students were significantly more empathetic ($t=2.75, p < 0.01$) during the second empathetic moment (M=4.10,SD=.85) than during the first empathetic moment (M=3.78,SD=.82). It is important to note that during the reflective self-assessment, the virtual tutor asked the students if they would change anything regarding their way to deal with Maria’s concerns, and 57% of the students indicated that they would not change their behavior. Additionally more than 80% of the students were rated higher or equal in the second empathetic opportunity as related to the first one.

### VII. Discussion

The primary result from the study was the correlation found between the self-assessment from students and the ratings provided by the three expert raters. While the correlation is not as strong as we originally expected, it is of significance. Visual inspection of the jitter plot shown in Figure 2 shows that there are a few examples of students over estimating their performance. Further inspection of these cases leads to the realization that these students might have interpreted the element of confirmation incorrectly. As noted before the opportunity presented to them was “Doctor, I don’t know if I’ll be able to pay for this treatment, are there other options?” The most common response to this opportunity was a direct answer to the question proposed by the patient, which included responses indicating the possibility for payment plans. While these responses acknowledge that Maria is concerned about the financial aspect of her condition, they do not necessarily deal with the anxiety expressed in her statement. Many students who answered with empathetic responses such as “We can setup a payment plan” also indicated during self-assessment that they confirmed her concerns. However, the third party
raters rated such responses as an implicit recognition of Maria’s concerns, which corresponds to a 2 in the ECCS scale. It is clear that in those cases, the design of the reflection could have been improved to create a better alignment between what the students and expert raters interpreted as confirmation. Additionally, an empathetic moment without the direct question could have also triggered more empathetic responses from the students. However, changing the empathetic moments was not possible due to the educational objectives of the experience. However, outside of those cases that limited themselves to answering the question proposed by the virtual patient, 41 students (70%) self-assessed themselves within one point as what the outside raters assessed. This behavior is encouraging for this kind of self-assessment, and we believe this confirms our hypothesis that this paradigm can lead to accurate self-assessment.

The secondary result shows that students were significantly more empathetic on the second empathetic opportunity than on the first, even when during their reflection they indicated they had no intention to change their response to Maria’s concerns. To further discuss this result we should look at one example. Participant 13 answered “Sure, we could extract the tooth” to Maria’s first empathetic question. It is clear that this was not a very empathetic response to Maria’s concerns. This particular participant proceed to rate herself as a 2 in the ECCS, meaning she implicitly recognized the concern but did not pursue the feelings being expressed. The expert raters agreed with this rating. Participant 13 also indicated that while her answer was not empathetic, she would not change the way she approached the empathetic opportunity provided to her. However, when presented with the second empathetic opportunity: ‘Doctor, Is this a long procedure? I had a bad dental experience when I was a child and have been fearful of going to the dentist since then.” her answer was “I will do my best to make you as comfortable as possible. I cannot give you a definite time line but I will try to work efficiently to provide you the best care possible. Tell me more about what happened when you were a child. I will try and avoid creating the same bad experience to the best of my ability. I’m so sorry you had a bad experience.” This answer was rated as a 5 by all 3 expert raters. While this is one of the few extreme cases of change, it was not an exception. This result seems to indicate that the explicit process of self-assessing their performance did create a significant positive impact in some of the students. These impact might not have been made explicit by the students, who mostly indicated they were not going to change their behavior. However, the implicit cognitive process did lead to significant differences. The significance of this result indicates that showing the students the components of a successful interaction during this in-action process of self-assessment can significantly improve future behavior. It is important to notice that this change happens even when both empathetic opportunities were rated as equally emotional by outside raters. However, both measures were taken just 5 minutes apart, so future research should look at the lasting effects of this behavior change.

VIII. CONCLUSION

We have introduced the concept of in-action reflection for self-assessment for virtual human experiences aimed at training interpersonal skills. This concept allows students to take their interpersonal skills training within this virtual human experiences one step further by giving them an opportunity to identify shortcomings in their performance, while providing them with a better idea of what an ideal interaction with the virtual human could be. This self-assessment is performed by rating different components of validated scales that measure the interpersonal skill the learner is training on. By performing this self-assessment through closed-ended questions, the student is presented with a roadmap that describes what they can do better in order to achieve better results. Indeed, the presented user study showed that even when students failed to indicate they were planning to change their behavior, the implicit learning through the self-assessment lead them to perform better in the second opportunity to demonstrate their interpersonal skills. Future research should look at the lasting effects of this behavior change we have observed.

While our study looked specifically at empathetic responses to a virtual patient, the concept of in-action reflections for self-assessment could be applied to other skills as long as a metric is identified for them. The metric would need to be adapted to close-ended questions that can be easily described to a user.

ACKNOWLEDGMENT

This research was partially supported by the National Institutes of Health (NIH) under award number 1R01LM010813-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES