Development of a Chatbot to Train Physiotherapy Students in Clinical Questioning and Reasoning

ChienChing Lee
Centre for Communication Skills
Singapore Institute of Technology
Singapore
chienching.lee@singaporetech.edu.sg

Malcolm Low Yoke Hean
Infocomm Technology
Singapore Institute of Technology
Singapore
malcolm.low@singaporetech.edu.sg

Benjamin Soon Tze Chin
Health and Social Sciences
Singapore Institute of Technology
Singapore
benjamin.soon@singaporetech.edu.sg

LiMing Lu
Infocomm Technology
Singapore Institute of Technology
Singapore
liming.lu@singaporetech.edu.sg

HweeHoon Lee
Centre for Communication Skills
Singapore Institute of Technology
Singapore
hweehoon.lee@singaporetech.edu.sg

Nadya Shaznay Patel
Business, Communication and Design
Singapore Institute of Technology
Singapore
nadya.patel@singaporetech.edu.sg

***Abstract*—This paper presents the development of a chatbot to train physiotherapy students in clinical questioning and reasoning. The features in this tool consist of semi-scripted practice conversations between a physiotherapist and patient to determine a clinical diagnosis on a given scenario, automated scoring and feedback on their performance, and instructor tracking of students' progress. Challenges and lessons learned in developing this chatbot are discussed. Experimental evaluation on whether the use of the chatbot improves students' self-efficacy in clinical questioning and reasoning is ongoing.**

Keywords—virtual patient, physiotherapy, semi-scripted practice, chatbot, Google Dialogflow

I. Introduction

The use of standardized patients (SP) in the clinical training of students has been found to benefit students. Standardized patients however, are costly, may not be able to depict a patient's condition realistically, and cannot cater to a large number of students [1]. Virtual Patient (VP) simulations overcome these issues and are gaining ground pre-Covid globally [2] and during this Covid period where social distancing measures are enforced. With VPs, students can practice using uniform and also rare but important clinical scenarios [3]-[5]. We see a big potential to tap on the use of chatbots to develop clinical cases which the allied health students can use to practice their history taking and communication skills. This paper aims to present the development process of a chatbot VP to improve physiotherapy (PT) students' clinical questioning and reasoning. Experimental evaluation on what kind of improvements can be seen in PT students' self-efficacy and empathy by using the chatbot is ongoing.

II. Related Work

A. *The use of Standardized Patients in Medical Training*

Bandura [6] defined self-efficacy as "the belief in one's capabilities to organize and execute courses of action required to produce given attainments" (p. 3). He further mentioned that self-efficacy impacts the learner's motivation and commitment to achieve their goals and persevere during challenges. Thus, learners who have higher levels of perceived self-efficacy are more likely to commit more effort to achieving their goals and reduce defensive behaviors.

The use of Standardized Patients (SP) in medical training is considered by scholars a near high fidelity simulation. This is because such simulation-based training is highly realistic since it involves human role players interacting with learners [7]. As an integral component for communication training, SPs are used in contexts that support both learning and assessment [8]. SPs can be actors, real patients or clinicians who are trained to follow predefined scenarios. They also provide standardized responses in response to interactions with learners from the perspective of many different roles [9]. However, there may be some limitations. Some studies report a certain level of inauthenticity to the use of SPs. As they are not real patients and would rely on their acting skills to depict various conditions and symptoms, SPs find it challenging [10]. Moreover, SPs may require a long training period to produce a high-quality simulation [11] and are costly to sustain [1]. In addition, as SPs are expected to provide feedback to learners amongst other tasks like keeping within the set context and script, the level of fidelity may suffer [12]. This may cause a negative learner experience by the learner [13]. Ensuring the accuracy and validity of an SP's performance is crucial [14]. Hence, it is crucial for medical training providers to ensure that performances by SPs are standardized equally so as to remove any measure of subjectivity.

B. *Training Students to be Empathetic*

Empathy training is considered by most medical schools as part of the curriculum relating to the subject of professional ethics [15]. Training in empathetic communication aims to teach medical students how to communicate with patients and their companions such that they are able to "walk a mile in their patients' shoes" [16]. Demonstrating empathy while engaging in clinical questioning is deemed challenging by learners. This is because as learners focus their questions to gather the most salient information in the clinical setting, empathy becomes less intentional [17]. Yet, without empathy, a medical student is arguably unable to carry out core medical tasks accurately as it is regarded as a key determinant of quality in medical care [18]. In fact, it is found to be associated with improved patient satisfaction, better diagnostic and clinical outcomes, and enhanced patient enablement [19]. However, scholars report that the skill to demonstrate empathy has appeared to decline during medical school and residency [20]. To make matters worse, instruments used to measure empathy in these studies were not validated with SPs' perceptions [21]. Furthermore, many studies argued that medical training in empathetic communication has been ineffectively conceptualized and operationalized [22].

C. *The Use of Chatbots in Medical Training*

Advanced technology in the form of virtual reality and AI chatbots has been used extensively for training and education purposes. Particularly in medical education, research has found value in the use of such technology to train history-taking as well as develop clinical reasoning and interaction skills [23]-[26]. For instance, VCAAI, an AI and a chatbot tool helps train nursing students in communication skills [27]; MediSIM, an AR tool, helps students see and feel the physical symptoms of the virtual patient [28]; and PASS-IT, a VR tool, helps medical students learn about patient safety [29]. AI chatbots, or conversational agents, can act and sound like real humans, and through audio and/or text, can simulate interaction with human users [30], thus potentially functioning well as virtual patients.

The proposed project will be a significant step forward in technology-enhanced learning for physiotherapy training in Singapore. Besides training in history taking and communication skills, a key aspect of the speech-enabled AI chatbot project is the inclusion of empathy expressions/ phrases in the grading system as an essential element in each student's conversation with the chatbot/virtual patient (VP). The use of a speech-enabled AI chatbot will provide a more authentic learning approach for students to formulate their clinical questions and reasoning skills during the history-taking process. The overall aim of the project is to improve the self-efficacy of the PT students via repeated practice and automated feedback in performing history taking and clinical reasoning with empathy using a speech-enabled AI chatbot.

III. SYSTEM ARCHITECTURE

Figure 1 shows an overview of the VP system architecture developed in this project. The backend chatbot logic of the VP is implemented using Google DialogFlow. The repository hosting the backend logging of students' input and the tracking of the scores achieved by each student is implemented using AWS Lambda and AWS S3 bucket. The instructors can access this repository through a web browser. The frontend 3D avatar that interacts with the students is implemented using Unity. Students can interact with the 3D avatar using text or voice input through a web browser.

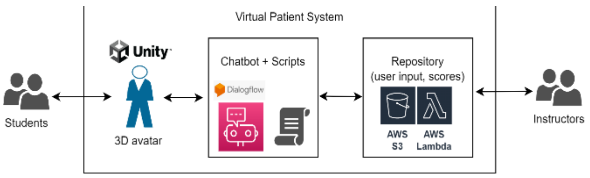

Fig. 1. Virtual patient system.

IV. MAIN FUNCTIONS OF THE SYSTEM

The features in this tool consist of semi-scripted practice conversations (audio or textual) between a PT and virtual patient (VP) in a musculoskeletal module to determine a clinical diagnosis using open-ended questions based on given scenarios and a practice checklist provided by the physiotherapy professor. To simulate a conversation with a patient, the students spoke to an avatar (the chatbot). Students could attempt each scenario multiple times. Upon completion of each attempt, students could view their score and a report so that they can self-evaluate and improve their performance. This is a proof-of-concept project which if found feasible, could be scaled up and improved in terms of accuracy and authenticity. Its use in the initial stages of development is mainly for practice and as a supplementary tool. The development team included the participation of three physiotherapy honors students and three ICT students, supervised by the respective professors.

A. *Management of the Practice Session*

The students access the chatbot for practice purposes using a unique URL sent to them, hosted in the university's learning management system. They then key in their SIT student ID to gain access to the practice sessions. Students can then choose a scenario that they would like to practice on, and the type of input preferred (audio via a microphone or text input where they type their questions in the textbox). Once the session is completed, the students could click on the 'view report' button to see their score and detailed report on their performance.

Students could practice their skills with the VP anytime, anywhere, repeatedly. Such flexibility and availability were enabled by the chatbot, which is not possible with SPs. The autonomy given to students lets them self-pace and self-learn in a non-threatening student-to-VP setting (as opposed to a clinical context), encouraging them to be independent yet accountable for their learning.

B. *Semi-scripted Practice Conversations*

The conversations between the PT and VP were pitched at year 1 students' level similar to that of their aural exam with standardized patients which required a direct response. It was felt that an actual clinical setting would need to take many more factors into consideration which the year 1 students were not ready for yet. The script for a scenario with a healthy and cooperative patient experiencing pain in the elbow was written by the physiotherapist student team. Empathy phrases following the Jefferson Scale of Empathy [31] were deliberately included into the conversations to raise students' awareness about the need to understand patients' pain besides coming to a clinical diagnosis.

The script was keyed into the DialogFlow platform which required two main components: intents and responses. Intents referred to predefined training phases. For example, the PT would begin a conversation with a 'greet' intent, followed by 'verify ID', 'patient query', 'empathy' and 'consent' intents. Each intent in the script was standardized to have five PT utterances and three responses by the chatbot. For example, the 'patient query' intent can have the utterances and responses as shown in Table 1. Any question that the PT asked (or typed) that matched any of the five utterances will be matched to the 'patient query' intent by DialogFlow, which will trigger any of the three responses to be spoken by the VP to the PT.

With multiple testing, more utterances covering all the sections in the checklist, including misspelling and mispronunciation, were added and utterances revised for brevity to facilitate more accurate recognition by the system.

TABLE 1. EXAMPLES OF UTTERANCES AND RESPONSES FOR THE 'PATIENT QUERY' INTENT

Utterances	Responses
• What can I do for you today? • How may I help you? • What brings you here today? • Why are you here today? • How can I help you today?	• I have pain in my right elbow. • My right elbow hurts. • My right elbow is painful and has been bothering me.

The script was also clearly demarcated by phases to facilitate reporting on students' performance.

The phases in the history taking conversation were designed referencing the Musculoskeletal Examination and Assessment framework [32]. The logical sequence of phases was specified by input tag and output tag of intents in DialogFlow. Each intent is associated with a set of input tags and output tags. An intent can only be matched by DialogFlow if all its input tags are present. The output tag determined the intents that could follow in the conversation. In our system, we used the numbered ID of an intent to label the input and output tags. Utterances that match intents of the input tags became contexts for the subsequent conversation. For example, "on a scale of 0 to 10, how painful do you feel?" would refer to the painful body part that was most recently mentioned. Further details on this implementation are explained in section V.B. Multiple intents could be enabled in the output tags, for a PT could ask multiple relevant next questions in a non-specific order. For example, suppose the 'greet' intent has an ID '1.1', the 'patient query' intent has an ID '1.2', and the 'empathy' intent has an ID '1.3'. The 'greet' intent has no input tag but has output tags '1.2' and '1.3'; the 'patient query' intent has input tag '1.2' and no output tag; and the 'empathy' intent has input tag '1.3' and no output tag. DialogFlow will not be able to match either the 'patient query' or the 'empathy' intent until the 'greet' intent is matched. After the 'greet' intent is matched, the output tags '1.2' and '1.3' are both generated. DialogFlow will now be able to match either the 'patient query' or the 'empathy' intent. A PT could also divert the conversation to an earlier but related phase, to supplement the history taking with skipped details. With multiple sample utterances and a flexible flow of the conversation phases, the semi-scripted system simulated the conversation that a PT would have in a clinical interview.

The conversation script was written at a moderate level of complexity, appropriate for the target audience. The moderate complexity of the task and the availability of a musculoskeletal framework is hoped to help students in interviewing VPs and boost students' self-efficacy in clinical interviews.

C. *Automated Feedback and Visualization*

Providing automated feedback for students to self-evaluate and improve their performance in each practice is important. The feedback was detailed as it displayed the correctly explored questions and the missed questions in two lists (Fig. 2). The positive feedback on correctly explored questions aims to affirm the students. Similarly, the feedback on the missed questions serves to guide the students to self-correct and to perform better in their next attempt. Over multiple practice sessions, both types of feedback serve as credible evidence to convince students themselves that they are improving, and that the progress was made through their efforts, thus strengthening the self-efficacy in their questioning and reasoning skills.

The students' score was based on a penalty system from a full score of 100 points (Fig. 3). Every missed question resulted in a deduction of one point while an empathy response was given a bonus one point as one of the aims of this project was to raise students' awareness regarding the need to be empathetic when treating patients. Yellow flag questions referred to questions that PT students needed to probe further while red flag questions referred to questions that cannot be missed as they were critical to clear potential serious underlying medical conditions pertaining to the case scenario. The AI chatbot included all the essential components of history taking [32] that is expected of students in their training.

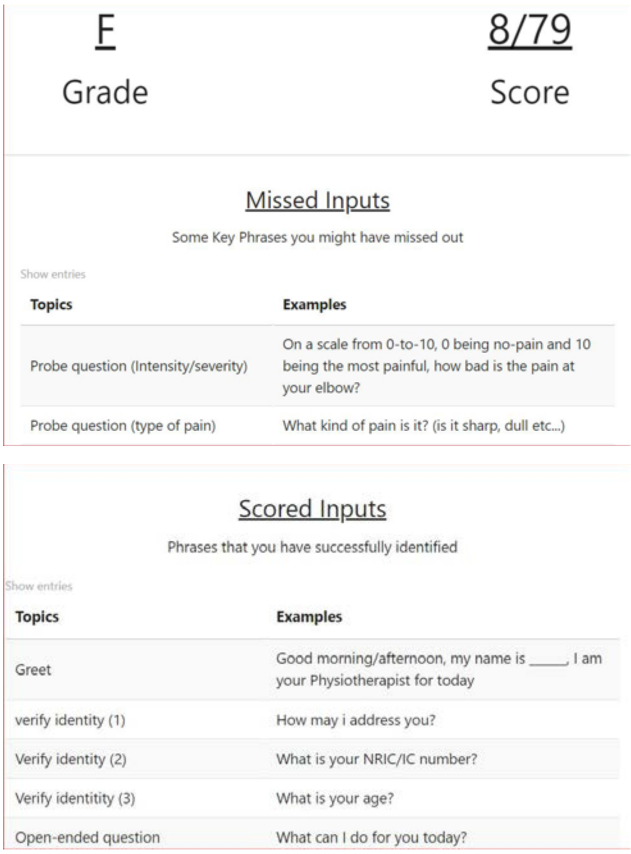

Fig. 2. Automated student feedback.

Marking scheme- Penalty system (%)	
Criteria	Points
Standard history taking questions	
Every missed question (input tags)	Minus 1 point
Empathy response	
Every tag	Add 1 bonus mark
Yellow and Red flag questions	
Every missed question (input tags)	Minus 2 points
Difficult question	
Every missed question (input tags)	Minus 1 point
Consent	
Every missed consent (input tags)	Minus 2 points

Fig. 3. Penalty system for scoring.

D. *Instructor View on Students' Performance*

The instructor can view students' performance via a downloaded excel file (Fig. 4). The scores were segmented into phases in the conversation so that instructors could identify the phase which students had the most missed questions and thus provide for further support to students. The instructor can also download an audio recording of the conversation between the PT and VP to analyze the questions asked by the students.

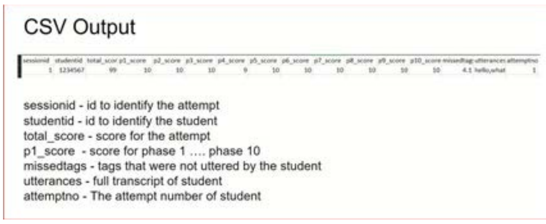

Fig. 4. Components tracked for students' performance.

V. CHALLENGES

During the implementation of the chatbot, the team faced several challenges in ensuring that the chatbot interacted with the students in a logical and meaningful way. Some of these challenges are explained in this section.

A. *Ensuring Logical Flow in the Script*

Our initial implementation of the chatbot allowed all the intents to be enabled right from the start of a consultation session. This allows students to skip over questions relating to greeting the patient and verifying the patient's identity and jump straight to asking questions relating to diagnosing the patient's medical condition. To mitigate this, we divided the set of questions into three stages. Stage 1 is the introduction phase where the PT will greet the VP and verify his or her identity. Stage 2 is the body phase where the PT will ask a set of questions to determine the type and nature of pain the VP is suffering from and the medical history of the VP. Stage 3 is the goal setting phase where the PT will ask a set of question to establish the long- and short-term goals of the VP. The input tag of questions in stages 2 and 3 will only be generated after some of the questions in the previous stage have been asked.

B. *Resolving Similar Questions Relating to Different Parts of the Body*

During the consultation with the VP, the PT may have to ask two or more very similar questions, but each referring to different parts of the body. For example, the PT may first ask the question "How bad is the pain on your **arm** on a scale of 0 to 10, where 0 is no pain, and 10 is the worst pain you can imagine". This will be matched to the chatbot intent **Intent-A** relating to the pain level of the patient's arm. Subsequently, a similar question "How bad is the pain on your **knee** on a scale of 0 to 10, where 0 is no pain, and 10 is the worst pain you can imagine" is asked by the PT. This question is supposed to be matched to the chatbot intent **Intent-K** relating to the pain level of the patient's knee.

However, during our testing, we found that both questions were matched to the same Intent-A relating to the pain level of the arm as the two utterances were very similar. To resolve this, we implemented specific input tags to each of the two intents. **Intent-A** was paired with **InputScaleTag-A** and **Intent-K** was paired with **InputScaleTag-K**. Initially, both InputScaleTag-A and InputScaleTag-K are disabled. Each of the input tag were enabled when the conversation with the chatbot entered the respective sections.

C. *Improving the Accuracy of Intent Matching*

To improve the accuracy of the matching of the PT's utterance to the correct intent, during the testing phase, our research assistants went through all the utterances that failed to match to any intent, and manually assigned them to the correct intents respectively. Over time, the accuracy of the matching improved. For example, the utterance "Can you tell me why you are here?", which did not match any intent initially, can be matched to the 'patient query' intent shown in Table 1 using the process above.

D. *Improving Conversational Engagement*

Currently, the PT is speaking to an avatar which is not able to portray the full range of emotions of a human being in pain. Thus, the PT may not experience more natural responses from the VP. Furthermore, the conversations are semi-scripted to enhance the accuracy of intent matching. These are limitations of the current chatbot which we hope to improve on in the future.

CONCLUSION

By giving PT students the opportunity to experience history-taking with VPs before their clinical placement and in class, they will have firsthand exposure to the clinical reasoning process in a safe, structured, and guided approach prior to them seeing actual patients. This is significant in light of limited placement resources available in public health institutions and the safety measures aligned with the COVID pandemic. We hope to present findings on whether our chatbot improves students' self-efficacy in clinical questioning and reasoning with empathy after we have completed our experimental evaluation.

ACKNOWLEDGMENT

This work is sponsored by the Applied Learning and Innovation Grant, Singapore Institute of Technology.

The authors would like to acknowledge the contributions of these Research Assistants in this project: Yeo Yue Heng, Abdul Hady bin Zaydie, Lau Hong Sheng, Sally Chua Yuann Tyng, Darren Soon Jia Jun and Gan Pei Li.

REFERENCES

[1] L. Bokken, J-J. Rethans, A. J. J. A. Scherpbier, and C. P. M. van der Vleuten, "Strengths and weaknesses of simulated and real patients in the teaching of skills to medical students: A review," Simul Healthc., vol. 3, pp. 161-169, June 2008.

[2] N. Al-Hiyari and S. Jusoh, "The current trends of virtual reality applications in medical education," in 12th International Conference on Electronics, Computers and Artificial Intelligence, Bucharest, Romania, June 25-27, 2020, pp. 1-6.

[3] C. D. Combs and P. F. Combs, "Emerging roles of virtual patients in the age of AI," AMA J. Ethics., vol. 21, E153-159, Feb. 2019.

[4] F. Consorti, R. Mancuso, M. Nocioni, and A. J. C. Piccolo, "Efficacy of virtual patients in medical education: A meta-analysis of randomized studies," Education, vol. 59, pp. 1001-1008, Nov. 2012.

[5] N. Posel, D. Fleiszer, and B.M. Shore, "12 tips: Guidelines for authoring virtual patient cases," Medical Teacher, vol. 31, pp. 701-708, Aug. 2009.

[6] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," Psychological Review, vol. 84, pp. 191-215, March 1997.

[7] F. Lateef, "The Standardized Patient in Asia: Are there unique considerations?" American Journal Emergency & Critical Care Medicine, vol. 3, pp. 031-035, May 2020.

[8] K. L. Lewis, C. A. Bohnert, and W. L. Gammon, “The association of standardized patient educators (ASPE) standards of best practice (SOBP),” Adv Simul, vol. 2, June 2017.

[9] P. A. Geoffroy, J. Delyon, and M. Strullu, “Standardized patients or conventional lecture for teaching communication skills to undergraduate medical students: a randomized controlled study,” Psychiatry Investig, vol. 17, pp. 299–305, Mar. 2020.

[10] P. R. Shankar and N. R. Dwivedi, “Standardized Patient's views about their role in the teaching-learning process of undergraduate basic Science medical students,” Journal of Clinical and Diagnostic Research, vol. 10, pp. JC01–JC5, June 2016.

[11] D. R. Xu, M. Hu, W. He, J. Liao, Y. Cai, S. Sylvia, K. Hanson, Y. Chen, J. Pan, Z. Zhou, N. Zhang, C. Tang, C., X. Wang, S. Rozelle, H. He, H. Wang, G. Chan, E. R. Melipillán, W. Zhou, and W. Gong, "Assessing the quality of primary healthcare in seven Chinese provinces with unannounced standardized patients: protocol of a cross-sectional survey,” BMJ Open, vol. 9, pp. 1-11, Feb. 2019.

[12] D. Papanagnou, M. R. Klein, and X. C. Zhang, "Developing standardized patient-based cases for communication training: lessons learned from training residents to communicate diagnostic uncertainty,” Adv. Simul, vol. 6, July 2021.

[13] E. T. Newlin-Canzone, M. W. Scerbo, G. Gliva-McConvey, and A. M. Wallace, “The cognitive demands of standardized patients: understanding limitations in attention and working memory with the decoding of nonverbal behavior during improvisations,” Simul Healthc, vol. 8, pp. 207-14, Aug. 2013.

[14] T. Rutherford-Hemming, C. M. Alfes, and T. L. Breymier, “A systematic review of the use of standardized patients as a simulation modality in nursing education,” Nurs Educ Perspect, vol. 40, pp. 84-90, Mar/Apr. 2019.

[15] J. Benbassat and R. Baumal, “What is empathy, and how can it be promoted during clinical clerkships?” Acad. Med, vol. 79, pp. 832–893, Sept. 2004.

[16] J. Shapiro, “Walking a mile in their patients’ shoes: empathy and othering in medical students’ education,” Philos, Ethics, and Humanit Med., vol. 3, p. 10, Mar. 2008.

[17] S. Haider, Q. Riaz, & R, Gill, “Empathy in clinical practice: a qualitative study of early medical practitioners and educators,” Journal of the Pakistan Medical Association, vol. 1, pp. 116-122, Jan. 2019.

[18] M. Neumann, J. Bensing, S. Mercer, N. Ernstmann, O. Ommen, and H. Pfaff, "Analyzing the "nature" and "specific effectiveness" of clinical empathy: a theoretical overview and contribution towards a theory-based research agenda,” Patient Educ Couns, vol. 74, pp. 339–346, Mar. 2009.

[19] F. Derksen, J. Bensing, and A. Lagro-Janssen, “Effectiveness of empathy in general practice: a systematic review,” Br J Gen Pract, vol. 63, pp. 76–84, Jan. 2003.

[20] M. Neumann, F. Edelh€auser, D. Tauschel, M. R. Fischer, M. Wirtz, C. Woopen, A. Haramati, and C. Scheffer, “Empathy decline and its reasons: a systematic review of studies with medical students and residents,” Acad Med, vol. 86, pp. 996–1009, Aug. 2011.

[21] J. A. Colliver, M. J. Conlee, S. J. Verhulst, and J. K. Dorsey, “Reports of the decline of empathy during medical education are greatly exaggerated: a reexamination of the research,” Acad Med, vol. 85, pp. 588–593, April 2010.

[22] S. H. Sulzer, N. W. Feinstein, and C. L. Wendland, “Assessing empathy development in medical education: a systematic review,” Med Educ, vol. 50, pp. 300–310, Mar. 2006.

[23] N.B. Berman, S.J. Durning, M.R. Fischer, S. Huwendiek, and M.M. Triola, “The role for virtual patients in the future of medical education,” Acad Med, vol. 91, pp.1217-1222, Sept. 2016.

[24] C.Y. Chang, G. J. Hwang, and M. L. Gau, “Mobile chatbot approach for nursing training,” Brit J. Ed. Tech, vol. 53, pp.171-188, 2022.

[25] M. Co, T. H. J. Yuen, and H. H. Cheung, “Using clinical history taking chatbot mobile app for clinical bedside teachings – a prospective case control study,” *Heliyon*, vol. 8, pp. e09751, June 2022.

[26] N. Saleh, “The value of virtual patients in medical education,” *Annals of Behavioral Science and Medical Education*, vol. 16, pp. 29–31, Oct. 2015.

[27] S. Shorey and E. D. Ng, “The use of virtual reality simulation among nursing students and registered nurses: a systematic review,” *Nurse Education Today*, vol. 98, pp. 104662, Mar. 2021.

[28] S. A. Cheow, “Using tech tools to replace real patients in medical school,” The Straits Times, Nov. 21, 2017. {online]. Available: https://www.straitstimes.com/singapore/health/using-tech-tools-to-replace-real-patients-in-medical-school, [Accessed Feb. 16, 2022].

[29] National University of Singapore, “Virtual reality gaming for medical students in the time of pandemic,” Aug. 5, 2020. [Online]. Available: https://medicine.nus.edu.sg/wp-content/uploads/2020/08/05Aug2020_PressRelease_-Patient-Safety-as-Interprofessional-Training-PASS-IT.pdf [Accessed June 29, 2022].

[30] A. Kaur, S. Singh, J. S. Chandan, T. Robbins, and V. Patel, “Qualitative exploration of digital chatbot use in medical education: A pilot study,” Digit Health, vol. 7, pp. 1–11, Aug. 2021.

[31] M. Hojat, J. DeSantis, S. C. Shannon, L. H. Mortensen, M. Speicher, L. Bragan, M. D. LaNoue, L. Calabrese, “The Jefferson Scale of Empathy: a nationwide study of measurement properties, underlying components, latent variable structure, and national norms in medical students,” Adv. In Health Sciences Ed. vol. 23, pp. 899-920, Dec. 2018.

[32] N. J. Petty and D. Ryder, Subjective Examination in Musculoskeletal Examination and Assessment. 5th edition. UK: Elsevier, 2017.