



# Evaluating Virtual Patient Interaction Fidelity With Advanced Communication Skills Learners

Stephanie Carnell<sup>1\*</sup>, Anna Miles<sup>2</sup> and Benjamin Lok<sup>3</sup>

<sup>1</sup>Department of Computer Science, University of Central Florida, Orlando, FL, United States, <sup>2</sup>Speech Science, The University of Auckland, Auckland, New Zealand, <sup>3</sup>Department of Computer and Information Sciences and Engineering, University of Florida, Gainesville, FL, United States

Previous research in educational medical simulation has drawn attention to the interplay between a simulation's fidelity and its educational effectiveness. As virtual patients (VPs) are increasingly used in medical simulations for education purposes, a focus on the relationship between virtual patients' fidelity and educational effectiveness should also be investigated. In this paper, we contribute to this investigation by evaluating the use of a virtual patient selection interface (in which learners interact with a virtual patient *via* a set of pre-defined choices) with advanced medical communication skills learners. To this end, we integrated virtual patient interviews into a graduate-level course for speech-language therapists over the course of 2 years. In the first cohort, students interacted with three VPs using only a chat interface. In the second cohort, students used both a chat interface and a selection interface to interact with the VPs. Our results suggest that these advanced learners view the selection interfaces as more appropriate for novice learners and that their communication behavior was not significantly affected by using the selection interface. Based on these results, we suggest that selection interfaces may be more appropriate for novice communication skills learners, but for applications in which selection interfaces are to be used with advanced learners, additional design research may be needed to best target these interfaces to advanced learners.

## OPEN ACCESS

### Edited by:

Salam Daher,  
New Jersey Institute of Technology,  
United States

### Reviewed by:

Deborah Richards,  
Macquarie University, Australia  
Laura Gonzalez,  
Universidad Argentina de la Empresa,  
Argentina

### \*Correspondence:

Stephanie Carnell  
stephanie.carnell@ucf.edu

### Specialty section:

This article was submitted to  
Virtual Reality in Medicine,  
a section of the journal  
Frontiers in Virtual Reality

**Received:** 25 October 2021

**Accepted:** 20 December 2021

**Published:** 10 January 2022

### Citation:

Carnell S, Miles A and Lok B (2022)  
Evaluating Virtual Patient Interaction  
Fidelity With Advanced  
Communication Skills Learners.  
*Front. Virtual Real.* 2:801793.  
doi: 10.3389/frvir.2021.801793

**Keywords:** virtual patients, medical simulation, virtual reality, simulation-based training, instructional design, cognitive load

## 1 INTRODUCTION

Virtual patients (VPs) are computer simulations of patients that allow healthcare students to practice a variety of clinical skills, ranging from physical exams to medical interviewing. Past work has established that VPs contribute to higher learning outcomes (Consorti et al., 2012; Cook et al., 2013; Hirumi et al., 2016). However, past research also notes that VPs—and more generally, simulation-based medical education—need more investigation regarding the interplay between a simulation's fidelity and its instructional effectiveness (Norman et al., 2012; Cook et al., 2013). For example, a 2012 article by Norman, Dore, and Grierson examined a number of studies that compared the learning effectiveness of high-fidelity simulators and low-fidelity simulators in medical education. Almost all of the twenty-four studies examined demonstrated that the high-fidelity simulations did not have a significant advantage over the low-fidelity simulations (Norman et al., 2012), despite what the authors note as the general assumption that higher fidelity simulations will yield better learning transfer to real-world scenarios.

This apparent conflict between simulation fidelity and learning effectiveness may stem from the *cognitive load* associated with higher fidelity simulations. Cognitive load theory explains the connections between working memory, the information a human can process consciously, and a learner's abilities to "...process new information and to construct knowledge in long term memory" (Sweller et al., 2019). Understanding the relationship between working memory and long-term memory is particularly important because working memory is limited, both in capacity and duration. In other words, if the cognitive load associated with a learning activity is too high, this load can impede the learning that may take place. Applied to simulation-based education, we may see the effect that high fidelity simulations may represent models that are too complex for novice learners (Norman et al., 2012), thereby reducing the potential educational effectiveness. To manage learners' cognitive load in simulations, some cognitive load researchers in medical education therefore suggest varying simulation fidelity in accordance with the expertise of the targeted learners: novice learners should use lower fidelity simulations, and as they learn, learners may progress to using simulations of gradually greater fidelity with adequate instructional support (Leppink and Duvivier, 2016). Similarly, the International Nursing Association for Clinical Simulation and Learning's best practices in simulation design suggests varying a simulation's fidelity in response to a number of factors, including the learner level and the learning objectives (Watts et al., 2021).

In this work, we apply this concept of "progressive fidelity" to VPs for medical communication skills training, specifically to the manner in which learners interact with a VP in a communication skills learning scenario. For example, learners may be able to chat "freely" with a VP using a chat interface, or learners may use a selection interface in which they select from a list of predetermined options to interact with the VP. Existing VP applications have used both chat and selection interaction methods, but the educational effectiveness of these methods and analysis of when they should be applied is under-investigated.

Our previous work examined fidelity and educational effectiveness in VP interaction methods by investigating the use of these interaction methods with learners of different experience levels (Carnell et al., 2015). Specifically, our work suggests that VP selection interfaces may be helpful for novice communication skills learners by providing examples of questions to ask in a patient interview. In this paper, we continue this investigation by evaluating the use of a VP selection interface with advanced medical communication skills learners. To perform this evaluation, we integrated VP interviews into a graduate-level course for speech-language therapists (SLTs) over the course of 2 years. In the first year, students interacted with three VPs using only a chat interface. In the second year, students used both a chat interface and a selection interface to interact with the VPs. Using survey responses from the learners and transcripts from their VP interactions, we investigated the following questions:

- When do advanced medical communication skills learners perceive that VP selection interfaces should be used in their learning?
- To what degree does conducting VP interviews using a selection interface impact advanced medical communication skills learners' communication behavior?

Our results suggest that 1) advanced learners perceive VP selection interfaces as more appropriate for novice learners and that 2) learners' communication skills behaviors were not impacted by the use of a selection interface. These results echo our previous findings that suggest the suitability of VP selection interfaces for novice learners. And so, for the case of using selection interfaces in VP communication skills training for advanced learners, we recommend that additional design research may be needed to target the selection interface appropriately to advanced learners. These results also reinforce the need for evaluation of VP development from educational perspectives, in addition to evaluation of the fidelity of the simulation, as the lower fidelity option (VP selection interfaces) may be an appropriate choice for certain educational contexts (training of novice communication skills learners).

## 2 RELATED WORK

In this section, we discuss existing applications that have used selection interfaces with virtual humans (VHs), the more general technology that VPs fall under. This discussion focuses on VH applications that have targeted communication skills learners' message production. We also discuss the issue of message production in the context of doctor-patient communication, the educational domain for which our VP selection interface was developed.

### 2.1 Existing Virtual Human Selection Interfaces

Existing VH applications targeting message production have used selection interfaces. For example, the Bilateral Negotiation Trainer (BiLAT) is a VH-centered game that focuses on intercultural communication and negotiation (Kim et al., 2009). During the game, the importance of message production and message reception (how one interprets a message) is emphasized. Players interact with VHs by selecting options from a menu, and VHs' responses may vary based on the level of trust a player has established. In an initial test of the learning effectiveness of BiLAT, results indicate that players without prior negotiation experience demonstrated significant improvement after using BiLAT (Hill Jr et al., 2006). However, for those players with negotiation experience, there was no significant difference in performance.

Another VH application that uses selection interfaces and that targets message production is SIDNIE. SIDNIE was developed to train nursing students to ask pediatric patients questions that are age-appropriate and unbiased (Dukes et al., 2013). SIDNIE

accomplishes this goal through a set of four learning activities, the majority of which involve using a selection interface to interview a virtual pediatric patient. Nursing students who used SIDNIE significantly improved their message production from the first activity to the last activity, but performance did not increase between each activity (Dukes et al., 2013). The second activity, a VP selection interview with immediate feedback, yielded a ceiling effect, while the third activity, grading an interview, yielded performance significantly worse than that of the previous activity.

In both BiLAT and SIDNIE, learner performance did improve as a result of using the application, but the exact effect of the selection interface on communication skills in practice has not been thoroughly investigated. Our previous research on VH selection interfaces, however, suggests that selection interfaces may impact message production: we compared the VP interview performance of healthcare students who had used a selection interface to the performance of healthcare students who did a different learning activity. Students who had used the selection interface asked VPs more questions that matched the system's natural language processing than students who completed a different learning activity (Carnell et al., 2015). This finding suggests that using a selection interface may impact a communication skills learner's behavior and that selection interfaces' effect on message production should be evaluated directly.

Message production improvement resulting from the use of SIDNIE and BiLAT was not uniform. Dukes et al. hypothesize that the lack of monotonic performance improvement between SIDNIE activities is due to the different levels of learning required in each of the activities, namely that grading an interview requires a higher level of learning than conducting an interview with a selection interface that provides immediate feedback (Dukes et al., 2013). This suggestion may also relate to the expertise-linked improvement in BiLAT, as the selection interface may have required a lower level of learning than was appropriate for the more advanced negotiators who participated. These results suggest that there are certain learning contexts—the expertise of the learner or the relative difficulty of the learning tasks—to which selection interfaces may be better suited than others.

We also previously explored the context in which to use selection interfaces, leading us to recommend the use of selection interfaces for novice medical interviewers in particular (Carnell et al., 2015). This recommendation resulted from student feedback that the students viewed the selection interface options as modeling opportunities, or examples of questions students could ask in a real patient interview. Our current work presented here investigates learner expertise and selection interfaces further by evaluating these interfaces with advanced learners.

## 2.2 Message Production in Doctor-Patient Communication

Problems in doctor-patient communication relating to message production are well-established. For example, in 1989, Bourhis, Roth, and MacQueen studied the medical language use of doctors, nurses, and patients (Bourhis et al., 1989). The authors described

doctors as “bilingual” since doctors must switch between everyday language and doctors' acquired medical language (Bourhis et al., 1989). Later research has demonstrated that message production problems persist in doctor-patient communication today. Koch-Weser et al. analyzed interactions between real doctors and patients and found significant differences in medical word use (Koch-Weser et al., 2009). In other studies, patients have also expressed concerns about healthcare providers who do not follow best practice communication strategies (Waisman et al., 2003; Shaw et al., 2009). In some cases, failure to practice these strategies has also led to malpractice lawsuits (Gordon, 1996).

These language problems can even persist after medical students receive communication skills training. Wouda and van de Wiel displayed this persistence in their work with medical students', residents', and consultants' communication skills (Wouda and van de Wiel, 2012). The authors found that new medical students performed as well as senior students on two sub-competencies: explaining and influencing. This similarity was present despite the senior students receiving training on explaining and influencing. As a result, VH developers should concern themselves not only with addressing message production in their scenarios but also addressing its effectiveness.

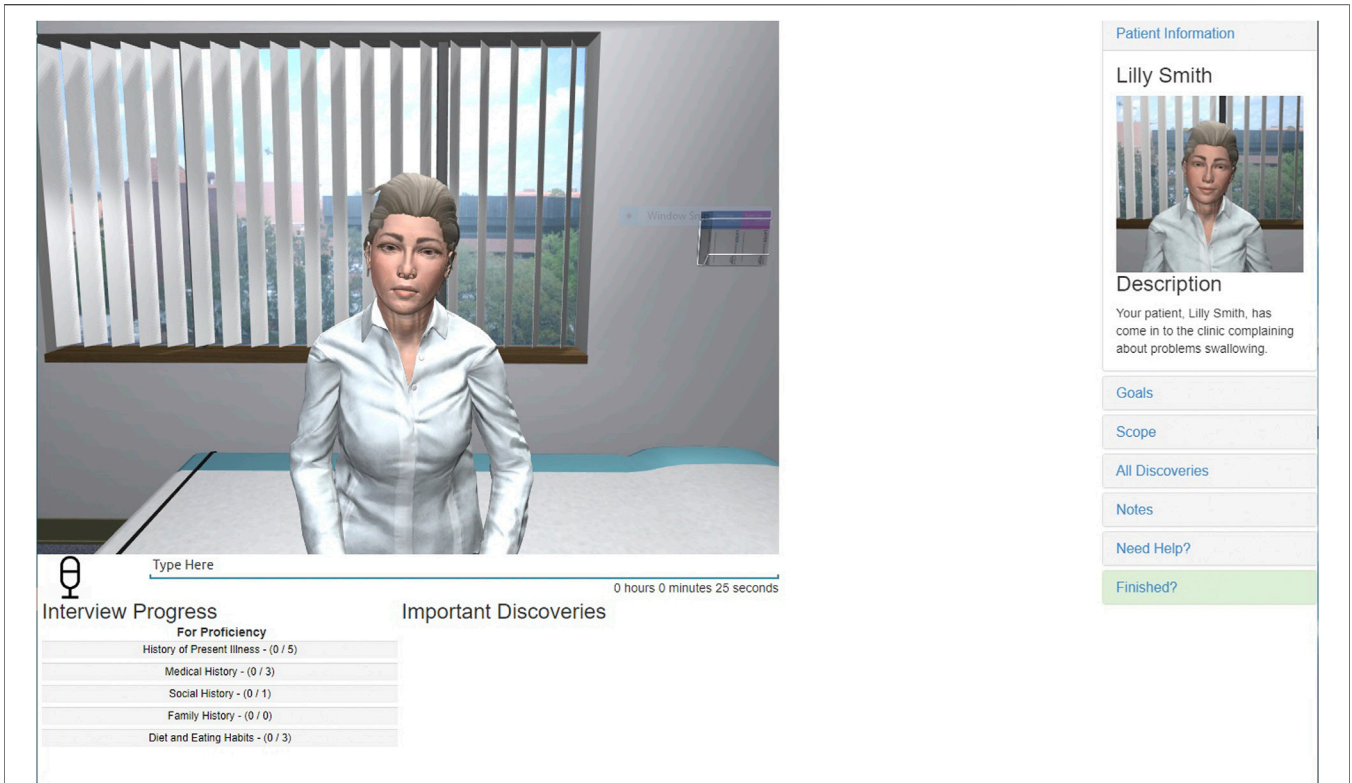
## 3 SYSTEM DESCRIPTION

For this work, we developed a new selection interface, the Guided Selection interface, to target message production. The Guided Selection interface was developed as part of an existing VP application, Virtual People Factory 2.0 (VPF2). In this section, we describe the features of VPF2 that are critical to the Guided Selection interface and the design of the new interface itself.

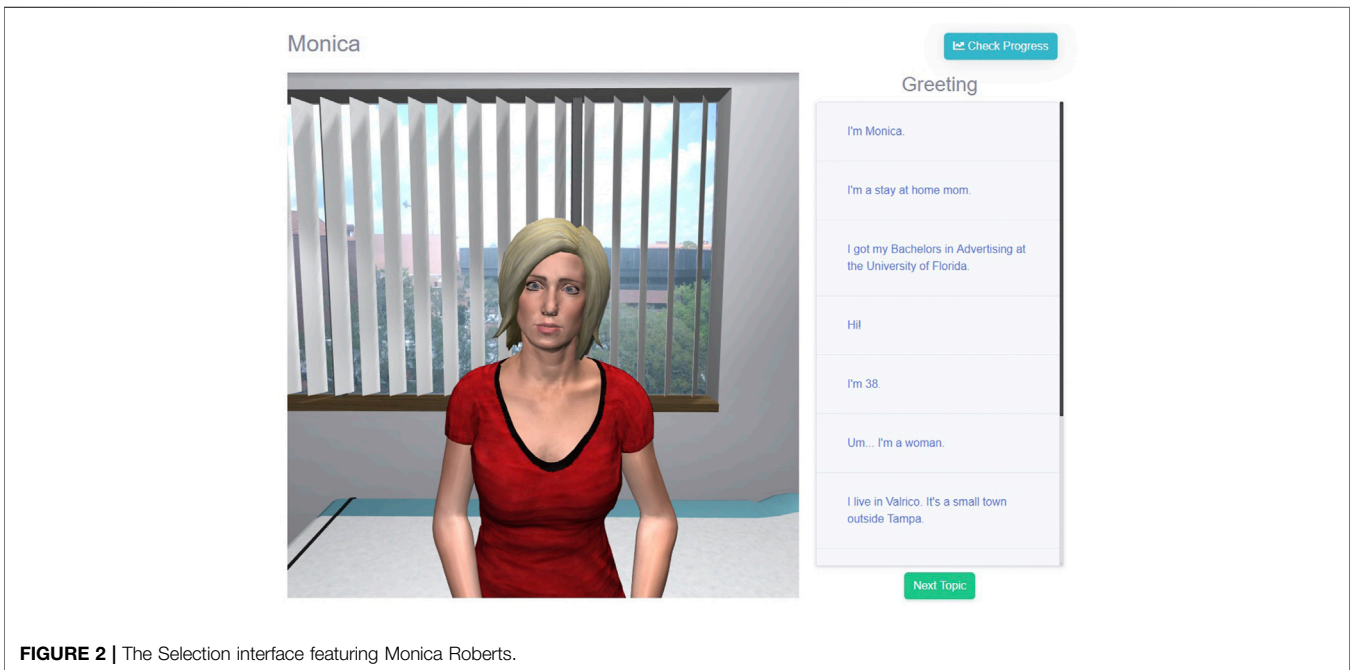
### 3.1 Virtual People Factory 2.0

Virtual People Factory 2.0 (VPF2) is a web application that enables the creation of and interaction with VHS (Rossen and Lok, 2012). VPF2 was designed to allow individuals without technical expertise but with particular domain expertise, such as a healthcare instructor, to author a VH that can then be interviewed in the same web application. The VPF2 authoring process mostly focuses on the creation of the VH script, which contains the dialogue responses a VH can provide, as well as the corresponding questions that can elicit those dialogue responses. VPF2 stores multiple questions that can elicit the same VH response. These multiple questions are called phrasings, as they are usually paraphrasings of the same question.

In addition to the VH authoring capabilities provided by VPF2, the application also enables the interviewing of VHS in an online interface. This interface allows remote learners to interview a VH using a personal desktop or laptop device. Typically, a VPF2 interview is conducted in a chat interaction style, as shown in **Figure 1**: learners type their questions into an input box, and VPF2 will match the typed question against the available phrasings in the VH script. If a matching phrasing is found, the VH responds with the corresponding VP response. If no matching phrasing is found, a standard exception response

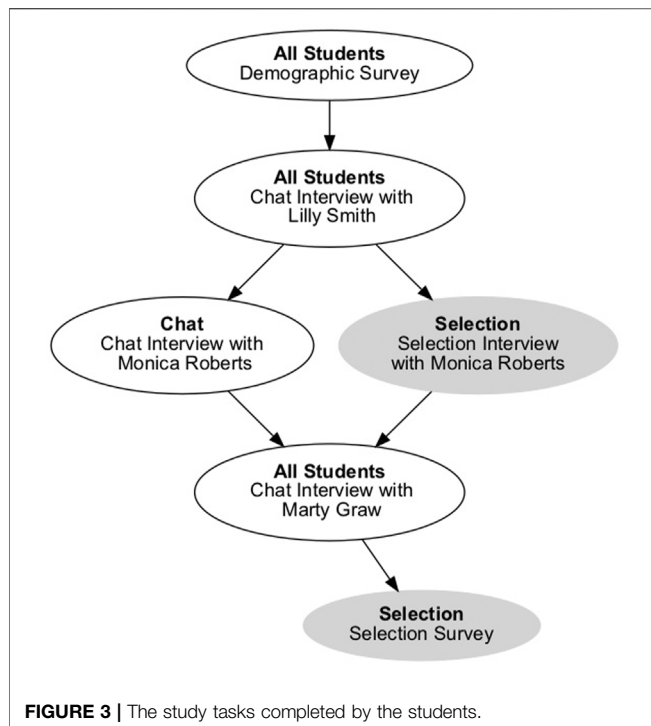


**FIGURE 1 |** The VPF2 chat interface with virtual patient Lilly Smith.



**FIGURE 2 |** The Selection interface featuring Monica Roberts.





“Sorry, I don’t understand what you just said. Can you say it another way?”) is returned instead.

### 3.2 The Guided Selection Interface

The Guided Selection interface (shown in **Figure 2**) can be used to conduct an interview on VPF2 by selecting questions from a predefined list. Questions are organized by the VP response they trigger. Every response in a VP’s script is listed with up to three phrasings that could trigger that response. The subset of phrasings displayed is determined by a numeric evaluation of the phrasings in the VP script. This numeric evaluation can be any relevant measure of language for the learning goal, but to target the healthcare students’ message production, the Flesch Reading ease (FRE) (Flesch, 1948) was used to measure a phrasing’s complexity, based on those issues identified in **Section 2.2**.

The FRE has been used in several studies to evaluate the language difficulty of patient-targeted health information (Bradshaw et al., 1975; Williamson and Martin, 2010; Agarwal et al., 2013). The FRE consists of a formula that uses a text’s words per sentence and syllables per word to calculate an overall score of reading difficulty for that text (Flesch, 1948). Higher FRE scores correspond to easier texts. Scores from the FRE typically range from 0 to 100, but scores outside of this range are also possible. Scores ranging from 60 to 70 are considered “standard” or “plain English” and correspond to a reading level for an American eighth or ninth grader (Flesch, 1949). For the Guided Selection interface, the FRE was calculated for each phrasing in the VP script. Using these FRE scores, up to three phrasings were selected for each VP response: the phrasings with the lowest, middle, and highest scores. If there were less than three phrasings for a VP

response, all of the phrasings were included. By providing learners with a variety of phrasings to elicit the same information, the Guided Selection interface may prompt learners to evaluate the complexity of their own questions in later interviews.

## 4 MATERIALS AND METHODS

To evaluate the Guided Selection interface’s impact on communication skills learners’ message production, we conducted a user study using two cohorts of a graduate-level, clinical practicum class at The University of Auckland. The user study’s main tasks were to conduct three interviews with VPs suffering from dysphagia, or trouble swallowing. An overview of these tasks is provided in **Figure 3**. Students enrolled in the clinical practicum class were well-suited to these tasks, as they were studying to be speech-language pathologists and would be required to interview real patients with dysphagia in their practice.

Students were recruited from two cohorts (2018 and 2019) of the clinical practicum class. All students interviewed the same VPs—Lilly Smith, Monica Roberts, and Marty Graw—in the same order. After each VP interview, students were asked to complete a diagnosis survey that asked about treatment plans and student concerns for the VP. One week after completing both the VP interview and the diagnosis survey, students were sent a link to a feedback page that summarized their performance with the VP. As the user study was integrated into a real class, there was roughly a month between each VP interview. Before completing any interviews, however, students were asked to complete a background survey that included information about their past experiences with medical interviewing and with relevant technologies.

Students always conducted their first and third interviews using VPF2’s chat interface (**Figure 3**). For the second interview, however, students used different interfaces based on the study condition they were in. For the CHAT condition, students continued to use the chat interface. For the SELECTION condition, students used the Guided Selection interface. Students were divided into conditions based on their class cohort: students in the 2018 cohort were in the CHAT condition, while students in the 2019 cohort were in the SELECTION condition.

### 4.1 Population

An overview of the SLP students who participated in the study is provided in **Table 1**. Twenty-four students completed the demographic survey and all three VP interviews (2018:  $n = 10$ , 2019:  $n = 14$ ). The average age of students in both conditions was  $28.0 \pm 9.11$  years, and the majority of the students (91.7%) identified as female. When asked to estimate how many real patients they had interacted with, students reported an average of  $29.9 \pm 15.5$  patients. Less than half of the students (41.7%) reported receiving any prior training for patient interviewing skills.

**TABLE 1** | Demographic information for the students who participated in the study.

	CHAT (2018 cohort)	SELECTION (2019 cohort)	Total
No. of Participants	10	14	24
Average Age (years)	27.4 ± 7.07	28.4 ± 10.6	28.0 ± 9.11
Female Participants	9 (90.0%)	13 (92.9%)	22 (91.7%)
Avg. Estimate of Patients Seen by Students	32.0 ± 16.2	28.4 ± 15.4	29.9 ± 15.5
No. of Students who Reported Receiving Interviewing Training	3 (30.0%)	7 (50.0%)	10 (41.7%)

## 4.2 Metrics

This user study used two sources of data: student survey responses and transcripts from students' interviews with the three virtual patients.

### 4.2.1 Student Survey Responses

For students in the SELECTION condition, we also asked them a series of questions about their experiences using the chat and Guided Selection interface to understand their perceptions of when each interface would be most useful. After their final VP interview, students in the SELECTION condition completed a survey comparing VPF2's chat interface with the Guided Selection interface. The survey included the following Likert scale items:

- I became frustrated while using the chat-based interaction method.
- I was able to ask the questions I wanted to using the chat-based interaction method.
- Using the chat-based interaction method helped me learn dysphagia interviewing skills significantly.
- Did the interview with the chat-based interaction method feel like a real-world interview? please explain why or why not.
- For what types of learning tasks do you think the chat-based interaction method is appropriate? For which tasks would it be inappropriate?
- What did you like or dislike about the chat-based interaction method?

These questions were then repeated for the Guided Selection interface by replacing "chat-based interaction method" with "selection-based interaction method." The questions were worded to be about the "interaction method" to direct students' attention to the manner in which they asked the questions, not the VP being interviewed. Before each set of questions, students were provided with a screenshot of the relevant interface and a brief description of when they used it. Items one through three were evaluated using a Likert scale, while the remaining questions were free-response. The first two Likert items were evaluated on a seven-point Likert scale, while the third item was evaluated on a five-point Likert scale.

### 4.2.2 Interview Transcripts

Transcripts of students' interviews with the VPs were analyzed to evaluate changes in students' message production. Literature in message production identifies three general categories of message

production assessment: goal attainment, efficiency, and social appropriateness (Berger, 2003). Based on these categories, we identified three message production metrics to evaluate the impact of the selection interface:

- Percentage of unique International Classification of Functioning, Disability, and Health Codes (Percent Unique ICF).
- Questions per discovery (QPD).
- Percentage of student utterances below the standard Flesch Reading ease (Percent Below Standard RE).

Transcripts were analyzed per student, per VP interview. While students were allowed to interview each VP multiple times, we chose only to analyze the transcripts with the longest time duration for each student to prevent unnecessary inflation of the metrics. For example, comparing questions per discovery across all of a student's transcripts may overestimate this measure by including all the questions a student asked in any interview.

Key to Berger's perspective of message production is the role of language as a tool to achieve a goal (Berger, 2003). According to the course instructor, an important goal when conducting an interview is to gain a holistic view of the patient. This holistic view can be achieved by asking questions that cover biomedical and social aspects of the patient's dysphagia. This goal, also known as patient-centered communication, is echoed in the medical communication literature: using a patient-centered approach has been shown to lead to improved patient satisfaction and cooperation (Smith et al., 1995; Beck et al., 2002), among other benefits.

To address students' use of patient-centered communication, we measured students' percentage of unique ICF codes (Percent Unique ICF). The World Health Organization's International Classification of Functioning, Disability, and Health (ICF) is a framework used to "describe and measure health and disability" (ICF, 2001). As part of this framework, the ICF includes a coding scheme to support a common vocabulary of health topics across disciplines and languages. Using the coding scheme provided by the ICF, we coded each student utterance to determine students' coverage of different health topics. Each question asked by the student was coded with one ICF code, and examples can be found in **Table 2**. For each student, we totaled the unique ICF codes used in each interview and normalized this count by the students' total utterances. A larger number of ICF codes used in a single interview likely indicates a more holistic view of the VP was pursued.

**TABLE 2** | Sample student utterances with information used to calculate goal attainment and social appropriateness metrics.

Student utterance	ICF code	Flesch reading ease
do you get a dry mouth	b5104 salivation	116.1
describe the sensation during swallowing	b51058 swallowing, other specified	15.64
do you work	d850 remunerative employment	119.2
How about physical activity?	d5701 managing diet and fitness	-8.725
can you feed yourself	e340 personal care providers	97.02
Are you having difficulty swallowing your medication	e1101 Drugs	6.359

The second category of message production assessment proposed by Berger is efficiency; speakers can potentially use multiple strategies to achieve their goals, but these strategies may vary in the time and effort needed (Berger, 2003). To measure efficiency, we measured students' questions per discovery. The questions per discovery metric originates from previous VP literature (Halan et al., 2018) and is the ratio of the number of questions asked by the student to the number of discoveries uncovered by the student. In VPF2, a discovery is an important piece of information needed to make a diagnosis. The questions per discovery metric reveals how efficiently a student can uncover important information in a VP interview. Higher values for questions per discovery indicate less efficient interviewing, as the student had to ask more number of questions to uncover discoveries.

Berger's final aspect of message production assessment is social appropriateness. As described in Section 2.2, the ability of a healthcare provider to adapt his or her message to a patient has been shown to be extremely important. Providers must communicate with patients in a manner that is comprehensible in order for patients to attend to providers' instructions. The two main suggestions given to providers to communicate in comprehensible language are 1) to speak in simple language (Graham and Brookey, 2008; Green et al., 2014; Speer, 2015) and 2) to use less medical jargon (Graham and Brookey, 2008; Oates and Paasche-Orlow, 2009; Green et al., 2014).

To target simple language, we calculated the percentage of student utterances below the standard reading ease (Percent Below Standard RE). This measure uses the Flesch Reading ease formula (FRE), the same readability formula used in the development of the Guided Selection interface. The FRE was calculated for each student utterance, examples of which can be found in Table 2. The percentage of student utterances below the standard reading ease addresses the general difficulty of a student's utterances by calculating the percentage of utterances that scored below 60, the lower end of the standard range of the FRE (Flesch, 1949).

## 5 RESULTS

### 5.1 Student Survey Responses

In contrast with Carnell et al.'s previous research investigating student perception of selection interfaces (Carnell et al., 2015), students in the SELECTION condition reported a higher level of

frustration when using the selection interface ( $5.64 \pm 1.34$ ) than the chat interface ( $4.79 \pm 0.802$ ). Students also reported a greater ability to ask the questions they wanted to ask using the chat interface ( $4.71 \pm 1.14$ ) than the selection interface ( $3.50 \pm 1.70$ ). Finally, students' responses indicate that the chat interface helped them learn dysphagia skills ( $3.29 \pm 0.611$ ) more than the selection interface did ( $2.86 \pm 0.770$ ).

There was some similarity with previous work and this current study in the themes that emerged from the open-ended survey questions. Select responses from each of the open-ended questions are presented in Table 3, Table 4 and Table 5. When describing which types of learning tasks the interfaces would be appropriate for, half of the students (7 students, 50.0%) described the selection interface as appropriate for learning what questions to ask or for learning about dysphagia in general, as shown in the example responses in Table 3. These tasks are likely novice tasks, given the students' status as advanced learners. In contrast, six of the students (42.9%) responded that the chat interface is appropriate for interview practice or taking a case history. These responses echo the previous finding that students view the selection interface as a modeling opportunity. They also suggest a progression of novice learning tasks (learning about dysphagia or learning questions) to more advanced tasks (taking a case history or practicing an interview) when using the two interfaces.

Students found both interfaces fairly unrealistic, as illustrated in the select quotes in Table 4: Only two students described the selection interface as realistic to some degree. Three students described the chat interface as realistic, and these responses were always qualified. For example, one student, Participant 11, responded: "Yes it was fairly realistic within the obvious restrictions of being an online client" - P11

Sample student responses as to whether students like or disliked a particular interface are included in Table 5. The lack of realism seemed to affect students' overall reception of the chat interface. The majority of students (10 students, 71.4%) reported disliking the chat interface or described it as "unnatural" in response to whether they liked the interface. Students also reported disliking the selection interface but for different reasons. Three students (21.4%) described using the selection interface as tedious, while two students (14.3%) felt they did not learn as much using the selection interface.

### 5.2 Interview Transcripts

To investigate whether advanced learners' message production was affected by using a selection interface, we conducted Mann-

**TABLE 3** | Sample student responses to “For what types of learning tasks do you think the [interface] is appropriate?”

Interface	Response
Selection	“For really structured interview practice where you don’t need to create questions . . . ”-P03
Selection	“ . . .Appropriate for introduction to different swallowing difficulties.”-P05
Selection	“appropriate for learning how to ask questions and evaluate our choice of wording against others. inappropriate for learning to identify gaps in information gathered”-P08
Chat	“appropriate: effective case history questions . . . ”-P20
Chat	“ . . .when gathering a case history this would be beneficial. Gives a guide to the conversation”-P10
Chat	“ . . .learning to have the discussion flow from one question to another”-P08

**TABLE 4** | Sample student response to “Did the interview with the [interface] feel like a real-world interview? please explain why or why not”.

Interface	Response
Selection	“ . . .I was not able to ask the questions I wanted to.”-P02
Selection	“ . . .In the real world you don’t get a list of set questions.”-P06
Selection	“yes, but again the same as the other interaction - the responses felt limited compared to real life.”-P20
Chat	“It felt robotic and artificial”-P07
Chat	“ . . .in real-life patients would respond more in an open-ended question . . . but these chat based interactions felt more limited than usual.”-P20

**TABLE 5** | Sample student response to “What did you like or dislike about the [interface]?”

Interface	Response
Selection	“I disliked it . . . you didn’t have to think through the series of questions and answers. . .”-P03
Selection	“ . . .I didn’t feel like I learnt anything.”-P06
Selection	“quick and simple. straightforward and no time wasted trying to re-word questions”-P20
Chat	“Disliked the interrogation feel”-P05
Chat	“ . . .I disliked that the questions/answers were limited.”-P08
Chat	“The unnatural pragmatic component mentioned above”-P10

**TABLE 6** | The means, standard deviations, and Mann-Whitney *U* test results for the three message production transcripts at Interview 1 and Interview 3. Metrics are defined in **Section 4.2.2**.

Interview 1	Mean (SD)			Mann-whitney U results		
	CHAT	Sel	U	<i>p</i>	<i>r</i>	
Message production metric						
Percent Unique ICF (%)	22.3 (6.08)	19.1 (4.89)	59.0	0.546	-0.275	
Questions per Discovery (QPD)	6.01 (1.45)	6.60 (2.75)	54.5	0.371	-0.185	
Percent Below Standard RE (%)	20.0 (7.34)	16.9 (4.76)	47.0	0.192	-0.132	

Interview 3	Mean (SD)			Mann-Whitney U Results		
	CHAT	SEL	U	<i>p</i>	<i>r</i>	
Message Production Metric						
Percent Unique ICF (%)	25.0 (8.06)	18.1 (5.32)	43.0	0.122	-0.324	
Questions per Discovery (QPD)	3.98 (1.33)	2.97 (0.970)	53.0	0.341	-0.204	
Percent Below Standard RE (%)	19.0 (8.19)	20.0 (9.16)	68.0	0.931	-0.0239	

Whitney *U* Tests on students’ message production metrics at Interview 1 and Interview 3.

During Interview 1, there were no significant differences in either message production metrics between students in the CHAT condition and students in the SELECTION condition. Similarly, during Interview 3, there were no significant differences between the two conditions in the two message production metrics. The medians, standard deviations, and the test results for the two metrics at Interview one and Interview three are presented in **Table 6**.

## 6 DISCUSSION

### 6.1 When Do Communication Skills Learners Perceive That Virtual Human Selection Interfaces Should be Used in Their Learning?

When asked which types of learning tasks the selection interface would be appropriate, one theme was tasks centered on learning,



such as learning about which questions to ask and how to ask these questions. Since the students in this work are advanced communication skills learners, these tasks are likely learning tasks associated with more novice learners. This theme of associating the selection interface with novice learners is consistent with feedback from novice students surveyed in our 2015 work (Carnell et al., 2015). Novice medical interviewers in the 2015 study described the selection interface as a resource for example questions and helpful for novice learners. Further, in both studies, students made these recommendations for the selection interfaces while acknowledging that selection interfaces are less realistic than chat interfaces. Describing selection interfaces as appropriate for novice learners despite a reduction in realism is consistent with recommendations made by Leppink and Duvivier to manage cognitive load in simulation-based medical education (Leppink and Duvivier, 2016).

However, one important distinction in this work is an additional type of novice-targeted activity proposed by advanced learners. While most feedback for the selection interfaces recommended novice practice for interviewing, three students described the selection interface as appropriate for “learning about dysphagia”. While learning about dysphagia was not specifically described as a novice activity by the students, since all of the students were at the end of their post-graduate-level training programs, we suspect they likely considered this to be a novice activity. Based on this recommendation from the students, we suggest that future research on the development of VP communication skills training may wish to examine how content instruction (such as that on dysphagia) may be integrated with communication skills practice for novices of both domains.

## 6.2 To What Degree Does Conducting Virtual Human Interactions Using a Selection Interface Impact Advanced Communication Skills Learners’ Message Production?

There were no significant differences observed in the three message production measures during Interview 3, so we do not find evidence that students’ message production was affected by the use of the selection interface.

This finding is interesting in light of our previous work which suggests that selection interfaces may serve as a modeling opportunity for questions to ask in a medical interview (Carnell et al., 2015). However, as noted, an important distinction between the 2015 study and this population is the participants’ levels of expertise. Since the participants in this work are considered advanced learners—in contrast to the novice learners previously studied—we hypothesize that the participants’ higher level of expertise may have mitigated potential modeling effects of the selection interface. This hypothesis is supported by the previous research with BiLAT, which found performance differences between novice and experienced negotiators who used BiLAT (Kim et al., 2009). In particular, the advanced learners in our study sought opportunities for autonomous communication with the VP by voicing a desire to “. . . to ask the questions I wanted to.”

Additionally, cognitive load theory may also explain why the selection interface did not impact learners’ message production, despite being viewed as a modeling opportunity. As described in **Section 1**, high fidelity simulations may be too cognitively demanding for novice learners to learn from them effectively. Cognitive load theory also suggests that the inverse may be true, that simulations that are too simplistic or of too low a fidelity may also be less effective for advanced learners. Therefore, if selection interfaces are better suited for novice learners, this may make selection interfaces less helpful for advanced learners. According to cognitive load theory, as learners become more advanced, the complexity of what can be stored in memory evolves. This evolution may make activities that are complex for novice learners too simple for advanced students. The presence of such an effect is suggested in the qualitative feedback given by students: they felt that they did not learn as much while using the selection interface, as shown in student responses to “What did like or dislike about the [interface]?”.

## 7 CONCLUSION

Previous research in simulation-based medical education has recommended greater investigation into the relationship between simulation fidelity and educational effectiveness, so this work examined advanced learners’ perceptions and use of interaction methods of varying fidelity for VP communication skills training. Cognitive load theory suggests that low fidelity simulations are often more suitable for novice learners, and our results suggest that advanced learners echo this suggestion: they recommended that the selection interface (the low fidelity interaction method) may be helpful for tasks such as learning which questions to ask and how to ask them. These tasks are likely novice tasks, given the learners’ status as post-graduate healthcare students. Further, in contrast to previous work that suggests VP selection interfaces may impact communication skills learners’ message production, advanced communication skills learners’ message production was not significantly different after using the selection interface. This finding may be potentially explained by cognitive load theory since a simulation may conversely be too simple for an advanced learner to benefit from it optimally. In summary, we present our conclusions below in reference to the questions we raised in **Section 1**:

- When do advanced medical communication skills learners perceive that VP selection interfaces should be used in their learning? Advanced communication skills learners seem to perceive selection interfaces as appropriate for novice learners [who are novices to medical interviewing but who are also potentially novices to dysphagia (see **Section 6.1**)].
- To what degree does conducting VP interviews using a selection interface impact advanced medical communication skills learners’ communication behavior? Advanced communication skills learners’ message production was not significantly different after using the selection interface proposed in this work. One potential

explanation for this finding comes from cognitive load theory (see **Section 6.2**), but further work will be needed to determine this.

## 7.1 Limitations and Future Work

A major limitation of our work is that we did not investigate directly whether the selection interface used in this study affected *novice* learners' message production. While our previous research suggests that novices' message production would be impacted (Carnell et al., 2015), we did not evaluate novices directly in this work. Future research should evaluate the same VP designs with learners of a variety of expertise levels simultaneously so as to best address the relationship between interaction fidelity and learner expertise level. This line of work is especially important, given learners' apparent perception that selection interfaces are most appropriate for novice learning activities.

These results are further limited by the amount of communication skills learners included in the study. We cannot conclude decisively from this study that selection interfaces are not effective at all for advanced communication skills learners, but this work does provide insight into advanced learners' perceptions of these interfaces. Based on our results, we suggest that if VP selection interfaces are to be used with advanced learners, additional design research is likely needed to target the selection interface to advanced learners. While our findings suggest that advanced learners do perceive selection interfaces as modeling opportunities, the other feedback from our learners indicates that they view these interfaces as appropriate for novice learners. Learners in this work were more frustrated using a selection interface than a chat interface and did not feel they learned interviewing skills as much with a selection interface as with a chat interface. Also, despite learners' describing the selection interface as a modeling opportunity, advanced learners who had used a selection interface

did not produce messages in a manner significantly different than learners who only used chat interfaces.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Florida Institutional Review Board, The University of Auckland Human Participants Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SC conducted the user studies, carried out the data analysis, and wrote the manuscript in consultation with AM and BL. AM was the instructor of record for the courses in which virtual patient interviews were integrated and chose the virtual patients to be interviewed. All authors contributed to the study design and conceptualization of the paper.

## ACKNOWLEDGMENTS

The authors thank Heng Yao for his role in running the user study and data collection in 2018, as well as the Virtual Experiences Research Group for their feedback and expertise in both VPF2 development and manuscript feedback.

## REFERENCES

- Agarwal, N., Hansberry, D. R., Sabourin, V., Tomei, K. L., and Prestigiacomo, C. J. (2013). A Comparative Analysis of the Quality of Patient Education Materials from Medical Specialties. *JAMA Intern. Med.* 173, 1257–1259. doi:10.1001/jamainternmed.2013.6060
- Beck, R. S., Daughtridge, R., and Sloane, P. D. (2002). Physician-Patient Communication in the Primary Care Office: a Systematic Review. *J. Am. Board Fam. Pract.* 15, 25–38.
- Berger, C. R. (2003). "Message Production Skill in Social Interaction," in *Handbook of Communication and Social Interaction Skills* (Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers), 257–289.
- Bourhis, R. Y., Roth, S., and MacQueen, G. (1989). Communication in the Hospital Setting: A Survey of Medical and Everyday Language Use Amongst Patients, Nurses and Doctors. *Soc. Sci. Med.* 28, 339–346. doi:10.1016/0277-9536(89)90035-x
- Bradshaw, P. W., Ley, P., Kinsey, J. A., and Bradshaw, J. (1975). Recall of Medical Advice: Comprehensibility and Specificity. *Br. J. Soc. Clin. Psychol.* 14, 55–62. Wiley Online Library. doi:10.1111/j.2044-8260.1975.tb00149.x
- Carnell, S., Halan, S., Crary, M., Madhavan, A., and Lok, B. (2015). "Adapting Virtual Patient Interviews for Interviewing Skills Training of Novice Healthcare Students," in *Intelligent Virtual Agents. Lecture Notes in Computer Science*. Editors W.-P. Brinkman, J. Broekens, and D. Heylen (Cham: Springer International Publishing), 9238, 50–59. doi:10.1007/978-3-319-21996-7\_5
- Consorti, F., Mancuso, R., Nocioni, M., and Piccolo, A. (2012). Efficacy of Virtual Patients in Medical Education: A Meta-Analysis of Randomized Studies. *Comput. Educ.* 59, 1001–1008. doi:10.1016/j.compedu.2012.04.017
- Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2013). Comparative Effectiveness of Instructional Design Features in Simulation-Based Education: Systematic Review and Meta-Analysis. *Med. Teach.* 35, e867–98. doi:10.3109/0142159X.2012.714886
- Dukes, L. C., Pence, T. B., Hodges, L. F., Meehan, N., and Johnson, A. (2013). "SIDNIE: Scaffolded Interviews Developed by Nurses in Education," in Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13, Santa Monica, California, USA, March 2013 (ACM Press), 395. doi:10.1145/2449396.2449447
- Flesch, R. (1948). A New Readability Yardstick. *J. Appl. Psychol.*, 32, 221–233. American Psychological Association. doi:10.1037/h0057532
- Flesch, R. F. (1949). *Art of Readable Writing*. New York, NY: Harper.
- Gordon, D. (1996). MDs' Failure to Use plain Language Can lead to the Courtroom. *CMAJ* 155, 1152–1154.
- Graham, S., and Brookey, J. (2008). Do Patients Understand? *Perm J.* 12, 67–69. doi:10.7812/tpp/07-144
- Green, J. A., Gonzaga, A. M., Cohen, E. D., and Spagnoletti, C. L. (2014). Addressing Health Literacy through clear Health Communication: A

- Training Program for Internal Medicine Residents. *Patient Educ. Couns.* 95, 76–82. doi:10.1016/j.pec.2014.01.004
- Halan, S., Sia, I., Miles, A., Crary, M., and Lok, B. (2018). “Engineering Social Agent Creation into an Opportunity for Interviewing and Interpersonal Skills Training,” in Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, Stockholm, Sweden, July 2018.
- Hill, R. W., Jr, Belanich, J., Lane, H. C., Core, M., Dixon, M., Forbell, E., et al. (2006). “Pedagogically Structured Game-Based Training: Development of the ELECT BiLAT Simulation,” in Proceedings of the 25th Army Science Conference., Orlando, FL, November 2006
- Hirumi, A., Kleinsmith, A., Johnsen, K., Kubovec, S., Eakins, M., Bogert, K., et al. (2016). Advancing Virtual Patient Simulations through Design Research and interPLAY: Part I: Design and Development. *Educ. Tech Res. Dev.* 64, 763–785. doi:10.1007/s11423-016-9429-6
- ICF (2001). *International Classification of Functioning, Disability and Health: ICF*. Geneva, Switzerland: World Health Organization.
- Kim, J. M., Hill, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., et al. (2009). BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. *Int. J. Artif. Intell. Educ.* 19, 21. doi:10.5555/1891970.1891973
- Koch-Weser, S., DeJong, W., and Rudd, R. E. (2009). Medical Word Use in Clinical Encounters. *Health Expect. : Int. J. Public Participation Health Care Health Pol.* 12, 371–382. doi:10.1111/j.1369-7625.2009.00555.x
- Leppink, J., and Duvivier, R. (2016). Twelve Tips for Medical Curriculum Design from a Cognitive Load Theory Perspective. *Med. Teach.* 38, 669–674. doi:10.3109/0142159X.2015.1132829
- Norman, G., Dore, K., and Grierson, L. (2012). The Minimal Relationship between Simulation Fidelity and Transfer of Learning. *Med. Educ.* 46, 636–647. doi:10.1111/j.1365-2923.2012.04243.x
- Oates, D. J., and Paasche-Orlow, M. K. (2009). Health Literacy: Communication Strategies to Improve Patient Comprehension of Cardiovascular Health. *Circulation* 119, 1049–1051. doi:10.1161/CIRCULATIONAHA.108.818468
- Rossen, B., and Lok, B. (2012). A Crowdsourcing Method to Develop Virtual Human Conversational Agents. *Int. J. Human-Comput. Stud.* 70, 301–319. doi:10.1016/j.ijhcs.2011.11.004
- Shaw, A., Ibrahim, S., Reid, F., Ussher, M., and Rowlands, G. (2009). Patients’ Perspectives of the Doctor-Patient Relationship and Information Giving across a Range of Literacy Levels. *Patient Educ. Couns.* 75, 114–120. doi:10.1016/j.pec.2008.09.026
- Smith, R. C., Lyles, J. S., Mettler, J. A., Marshall, A. A., Van Egeren, L. F., Stoffelmayr, B. E., et al. (1995). A Strategy for Improving Patient Satisfaction by the Intensive Training of Residents in Psychosocial Medicine. *Acad. Med.* 70, 729–732. doi:10.1097/00001888-199508000-00019
- Speer, M. (2015). “Using Communication to Improve Patient Adherence,” in *Communicating with Pediatric Patients and Their Families: The Texas Children’s Hospital Guide for Physicians, Nurses and Other Healthcare Professionals* (Houston, USA: Texas Children’s Hospital), 221–227.
- Sweller, J., van Merriënboer, J. J. G., and Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* 31, 261–292. doi:10.1007/s10648-019-09465-5
- Waisman, Y., Siegal, N., Chemo, M., Siegal, G., Amir, L., Blachar, Y., et al. (2003). Do Parents Understand Emergency Department Discharge Instructions? A Survey Analysis. *Isr. Med. Assoc. J.* 5, 567–570.
- Watts, P. I., McDermott, D. S., Alinier, G., Charnetski, M., Ludlow, J., Horsley, E., et al. (2021). Healthcare Simulation Standards of Best Practice™ Simulation Design. *Clin. Simulation Nurs.* 58, 14–21. doi:10.1016/j.ecns.2021.08.009
- Williamson, J. M. L., and Martin, A. G. (2010). Analysis of Patient Information Leaflets provided by a District General Hospital by the Flesch and Flesch-Kincaid Method. *Int. J. Clin. Pract.* 64, 1824–1831. doi:10.1111/j.1742-1241.2010.02408.x
- Wouda, J. C., and van de Wiel, H. B. M. (2012). The Communication Competency of Medical Students, Residents and Consultants. *Patient Educ. Couns.* 86, 57–62. doi:10.1016/j.pec.2011.03.011

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Carnell, Miles and Lok. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.