

# Initial construct validity evidence of a virtual human application for competency assessment in breaking bad news to a cancer patient

Timothy C Guetterman<sup>1</sup>  
 Frederick W Kron<sup>1</sup>  
 Toby C Campbell<sup>2</sup>  
 Mark W Scerbo<sup>3</sup>  
 Amy B Zelenski<sup>4</sup>  
 James F Cleary<sup>5</sup>  
 Michael D Fetters<sup>1</sup>

<sup>1</sup>Department of Family Medicine,  
 University of Michigan, Ann Arbor, MI,

<sup>2</sup>Department of Medicine, University  
 of Wisconsin–Madison, Madison,  
 WI, <sup>3</sup>Department of Psychology, Old  
 Dominion University, Norfolk, VA,

<sup>4</sup>Department of General Internal  
 Medicine, University of Wisconsin–  
 Madison, University of Wisconsin Medical  
 Foundation, <sup>5</sup>Department of Medicine,  
 School of Medicine and Public Health,  
 University of Wisconsin–Madison, Clinical  
 Science Center, Madison, WI, USA

→ Video abstract



Point your Smartphone at the code above. If you have a  
 QR code reader the video abstract will appear. Or use:  
<http://youtu.be/0AKsLrG7-Tk>

Correspondence: Timothy C Guetterman  
 Department of Family Medicine, University  
 of Michigan, 1018 Fuller Street, Ann Arbor,  
 MI 48104-1213, USA  
 Tel +1 734 998 7120  
 Email [tguetter@umich.edu](mailto:tguetter@umich.edu)

**Background:** Despite interest in using virtual humans (VHs) for assessing health care communication, evidence of validity is limited. We evaluated the validity of a VH application, MPathic-VR, for assessing performance-based competence in breaking bad news (BBN) to a VH patient.

**Methods:** We used a two-group quasi-experimental design, with residents participating in a 3-hour seminar on BBN. Group A (n=15) completed the VH simulation before and after the seminar, and Group B (n=12) completed the VH simulation only after the BBN seminar to avoid the possibility that testing alone affected performance. Pre- and postseminar differences for Group A were analyzed with a paired *t*-test, and comparisons between Groups A and B were analyzed with an independent *t*-test.

**Results:** Compared to the preseminar result, Group A's postseminar scores improved significantly, indicating that the VH program was sensitive to differences in assessing performance-based competence in BBN. Postseminar scores of Group A and Group B were not significantly different, indicating that both groups performed similarly on the VH program.

**Conclusion:** Improved pre–post scores demonstrate acquisition of skills in BBN to a VH patient. Pretest sensitization did not appear to influence posttest assessment. These results provide initial construct validity evidence that the VH program is effective for assessing BBN performance-based communication competence.

**Keywords:** verbal behavior, health communication, informatics, clinical competence, empathy

## Background

Effective communication is arguably the most important professional skill that a physician can possess.<sup>1–4</sup> However, training physicians in crucial interpersonal and communication skills remains a challenging problem for medical educators.<sup>5</sup> An often-overlooked aspect of communication training is assessment of learners' communication competency. Although coaching, communication workshops, and other techniques are sometimes used to train communication competency in medical education, teaching and assessing communication skills customarily occurs through the use of SPIs as part of OSCEs.<sup>6–8</sup> Developed in the 1960s,<sup>9</sup> SPIs have proven adept at assessing learners' "nontechnical" skills,<sup>7,8</sup> provided that SPI programs provide ongoing quality assurance monitoring to ensure standardized administration and reliability. Using SPIs for assessing communication, professionalism, and interpersonal interactions, however, presents some substantial problems.<sup>10–14</sup> SPIs are prone to fatigue and excessive mental workload, which limits their ability to correctly identify and report on critical conversational and behavioral cues.<sup>15</sup> They lack voluntary control over nonverbal behaviors,



so their communication can seem inauthentic and potentially confusing to learners.<sup>16,17</sup> Comprehensive assessment of large numbers of students, one-to-one, raises logistical problems related to time, cost, and room availability. Other concerns include the validity of assessment, which relies on examiner expertise and requires consistency across examiners.<sup>18,19</sup>

One solution for reliable and consistent assessment of communication skills is to use novel, computer-based methods, such as VH patients.<sup>20–22</sup> VHs are “intelligent” computer-generated, conversational agents with human form and the ability to interact with humans using verbal and nonverbal behaviors very similar to those people use in face-to-face interactions with each other. Researchers have developed and deployed successful military applications using VHs.<sup>23,24</sup> VH simulation also appeals to medical and nursing students, who are enthusiastic about learning using new media technology, which can promote engagement and enhance learning.<sup>25–27</sup> Preliminary efforts in the use of VHs to train medical communication and interpersonal skills have been exploratory.<sup>20,28</sup> Research on VH technology’s effectiveness as an assessment tool for clinical skills has been limited to examining the outcomes of 3-D VH patients as a component in a virtual OSCE compared to a traditional OSCE.<sup>29</sup>

One content area that presents a good fit for VH communication assessment and training is the field of cancer care. First, the need for communication training is clear. Conversations around cancer care are particularly stressful for physicians, patients, and family. When poorly executed, these conversations can potentially harm the peace and dignity of human beings when they are extremely vulnerable.<sup>30–36</sup> Despite its importance, only 5% of practicing oncologists report having been trained in basic communication skills such as relaying bad news,<sup>37</sup> and only 31.2% of terminally ill cancer patients report having had end-of-life discussions with their physicians.<sup>38</sup> Second, educational protocols already exist to provide communication training in cancer care.<sup>39,40</sup> These protocols are amenable to adaptation into virtual training environments and consequently provide a structure of educational domains to guide assessment.

Aware of the important need and the existing educational protocols, and utilizing grant funding from the National Institutes of Health, National Cancer Institute, three of the authors of this study (FWK, MWS, and MDF) – in cooperation with others – successfully developed and tested an innovative computer application MPathic-VR (an acronym derived from the grant title, “Modeling professional attitudes and teaching

humanistic communication in virtual reality”), which uses emotive VHs to assess and train communication skills in the setting of cancer. The initial step in MPathic-VR involved learning how to break bad news to a young woman (a VH) with leukemia. Our aim for this study was to gather evidence of construct validity for MPathic-VR’s particular use, namely, assessing physician performance-based competence in BBN. The construct being measured is communication competence upon exposure to a BBN module.

For the purpose of this study, we adopted Messick’s<sup>41</sup> highly cited unified concept of construct validity that subsumes content-related, substantive, structural, generalizable, external, and consequential aspects. At least four types of studies might be conducted to establish evidence of construct validity: 1) correlational studies, 2) multitrait–multimethod studies, 3) factor analytic studies, and 4) group difference studies, such as the one we have conducted. Our initial effort to gather construct validity evidence began by using an experiment to see whether MPathic-VR could detect pre–post differences in a group exposed to a communication training intervention. Although we considered a factor analysis or correlational study, as Messick<sup>41</sup> noted, “Probably even more illuminating in regard to score meaning are studies of expected performance differences over time, across groups and settings, and in response to experimental treatments and manipulations”. Hence, we chose the fourth type, namely, group difference studies.

## Methods

### Design

We used a two-group quasi-experimental design with internal medicine residents participating in a 3-hour seminar on BBN to cancer patients. To control for pretest sensitization, Group A completed the VH simulation before and after BBN seminar exposure, while Group B completed the VH simulation only after the BBN seminar.

### Participants

Second-year and third-year internal medicine residents at the University of Wisconsin in the Midwestern United States participated in a seminar, WiTalk, on BBN to patients, who were the target audience. The Group A seminar included 15 residents. Twelve days later, Group B attended the seminar, which included 12 residents. Seven residents (Group A,  $n=4$ ; Group B,  $n=3$ ) had previous exposure to a BBN training. The primary data collection occurred in the period January–February 2011.

## Ethics approval and consent to participate

The study received IRB exemption from the University of Wisconsin–Madison under the exemption category of research involving the use of educational tests. The research adhered to IRB guidelines for human subject protection. All residents provided informed consent to participate in the research.

## Educational intervention

The half-day seminar, called WiTalk (now presented in a substantially expanded format called WeTalk) was developed by two of the authors (TCC and ABZ), based on published reports and previous experience participating in the Oncotalk program.<sup>37,39,40,42,43</sup> A primary goal of WiTalk was for individuals to learn the SPIKES protocol for BBN.<sup>40</sup> SPIKES provides a structured method for BBN, in which the physician attends to “setting up” the meeting, assesses the patient’s “perception” of what is occurring, obtains an “invitation” to provide information, provides “knowledge” information to the patients about their condition, uses empathic responses to address “emotions”, and then “summarizes” the conversation and the strategy.<sup>40</sup> The half-day WiTalk workshop included the following core components: a brief didactic period of instruction practicing the skills, critiquing a colleague, and playing the role of a physician using standardized patient actors about realistic internal medicine cases in a small group of three to five people. Finally, trained faculty led an exercise to develop individual future learning objectives.

## Assessment of the intervention using the interactive VH program

The interactive program consisted of a large video monitor, a webcam, a microphone, and a Windows desktop computer loaded with the MPathic-VR program. The scenario began with a brief, introductory sequence that set up an ER encounter between the participant and a young VH woman named Robin. The sequence began with a handoff between the physician going off shift and the participant. The participant finds out that Robin is a previously healthy female who presented to the ER with a nosebleed. Her nose was packed, which controlled the bleeding. She is very impatient and would like to leave although her laboratory tests, including a CBC, are pending. The participant receives information through a call from the hematopathologist and learns that Robin’s platelets are dangerously low and that there are blasts with Auer rods on her blood smear, and the results are consistent with a diagnosis of acute myelogenous leukemia.

The learner must break this bad news to Robin and, in part through building rapport and demonstrating empathy, convince her to be admitted for further urgent evaluation and treatment. A demonstration video of MPathic-VR is available as a [supplementary video file](#).

The MPathic-VR program allowed participants to navigate the scenario by engaging in a conversation with Robin. The iteration of the MPathic-VR application used in this study featured a series of 14 key communication interchanges strung together to create a learner–VH dialog. In each interchange, the MPathic-VR application presented the learner with three possible responses to speak to Robin, who in turn then spoke with the learner. Participants were instructed that, “Your task is to pick what you believe to be the most appropriate statement from the set of three, and then speak it to the patient.” Each interchange included one optimal response and two suboptimal responses, which were plausible distractors, developed through work with cancer care experts (JFC and TCC) who have substantial experience in communicating bad news in cancer care. To ensure content relatedness, they reviewed the text for each choice, ranking of choices, and penalty values. Participants were penalized points for choosing suboptimal responses. The optimal choice, suboptimal choices, and the severity of penalties were based upon best practices for using the SPIKES protocol in BBN. Most penalty values range from 1.0 to 3.0. Higher point values are assigned to more serious deviations from protocol or good practice. Thus, a higher overall score reflects worse performance, and a lower score reflects better performance. The object of the interaction was to share the presumed diagnosis with Robin in an optimal manner and to transition her to the appropriate inpatient care. In addition to penalty points, suboptimal communication could also result in Robin leaving the ER in a very precarious condition. Table 1 provides a brief excerpt of the introductory script, demonstrating the statement–three response structure of an exchange. MPathic-VR uses data from each interchange (ie, response spoken by the learner) in combination with the programmed algorithm to determine the real-time VH response. For example, an optimal response by the learner leads to an appropriate and more neutral VH response, while a suboptimal response can escalate the interaction.

## Experiment

Group A completed the VH simulation prior to and after exposure to the BBN seminar to produce pre- and postseminar scores. It is possible that performance on the preseminar test could provide information and cues that might benefit

**Table 1** Excerpt from MPathic-VR breaking bad news script**Exchange 6****Robin**

Am I anemic? You know I'm a vegan, right? I promise I'll do better about taking my B-12. (annoyed)  
Can I go now?

**Doctor**

- (a) No, the problem is not anemia, and B-12 won't help
- (b) Robin, you have leukemia
- (c) Now, please listen to me carefully. I will answer all of your questions afterward

**Notes:** Each choice is distinct. The choice made contributes to the VH's reaction and participant scoring.

**Abbreviation:** VH, virtual human.

participants on the postseminar test. Thus, to account for the potential of testing as a validity threat, an additional group was needed. Group B attended the BBN training, but only participated in the simulation after the seminar, thus providing only postseminar test scores. If the training was effective, it was expected that the postseminar test scores would be comparable for Groups A and B. Because Group B did not complete a preseminar test, their postseminar test scores were free of potential pretest sensitization.<sup>44</sup> The dependent variable was the score on the MPathic-VR simulation.

## Hypotheses

We proposed the following hypotheses: 1) MPathic-VR scores will improve (decreased score reflects better performance) from the preseminar test to the postseminar test based on exposure to the BBN intervention; 2) improvement in scores results from improved understanding of seminar content, not pretesting; and 3) the difference will be greatest among individuals who had no BBN training before participating in the seminar.

## Analysis

Descriptive data were calculated to compare the two groups. We assessed pretest group differences to questions about experience in BBN using a Mann–Whitney *U*-test. Pre- and postseminar differences for Group A were analyzed with a paired *t*-test. The postseminar scores of Groups A and B were analyzed with an independent *t*-test. Differences between Group A's preseminar result and Group B's postseminar result were analyzed with an independent, directional (one-tailed) *t*-test. These data were analyzed with a directional *t*-test because the postseminar scores of Group B were expected to show improvement relative to the preseminar scores of Group A. A significance level of  $p < 0.05$  was the criterion for all statistical testing. All analyses were performed using

IBM SPSS Statistics, version 22 (IBM Corporation, Armonk, NY, USA).

## Results

The demographics of Groups A and B appear in Table 2. In Group A, which used the MPathic-VR program along with both pre- and postintervention tests, all participants were in their second year of residency. In Group B, which only used MPathic-VR postseminar test, seven participants were second year residents, three were third year residents, and two were unknown. Experience with BBN did not differ significantly between the groups, based on the results of the Mann–Whitney *U*-test.

Table 3 presents the mean MPathic-VR scores for both groups. Group A consisted of 15 participants, including four individuals with prior BBN training. Group B had 12 participants, including three with prior BBN training. Data from both groups met normality assumptions, based on the Shapiro–Wilk test ( $W=0.948$ ,  $p=0.49$ ;  $W=0.906$ ,  $p=0.12$ , respectively). For Group A, the postseminar mean score of 8.1 (SE: 0.69) was significantly lower than the preseminar mean score of 12.7 (SE: 1.24;  $t(14) = 3.41$ ,  $p=0.002$ ), demonstrating responsiveness to the intervention. These results support the hypothesis that scores improved pre–post as a result of exposure to the BBN seminar.

The postseminar test scores for Groups A and B did not differ significantly, indicating that both groups had a similar performance in MPathic-VR. As shown in Table 3, Group A scored a postseminar mean of 8.1 (SE = 0.7), and Group B scored a postseminar mean of 10.3 (SE = 1.2;  $t(16) = -1.4$ ,  $p=0.175$ ). Levene's test for this comparison was significant ( $p < 0.05$ ), so the *t*-value and *df* are reported for unequal variances.

The mean scores for the 12 participants in Group A, who reported no prior BBN training, are shown in Table 4. Posttest scores (mean = 8.6, SE = 0.80) were significantly lower (as noted, lower is better) than the pretest scores (mean = 14.0, SE = 1.29;  $t(11) = 3.32$ ,  $p < 0.007$ ). Further, we expected the posttest scores for Group B to be better (ie, lower) than the pretest scores of Group A. The results confirm that the Group B posttest scores (mean = 10.1, SE = 1.64) were significantly lower than the Group A pretest scores (mean = 14.0, SE = 1.29;  $t(19) = 1.9$ ,  $p=0.037$ ; Levene's test:  $p > 0.05$ ). The posttest means for both groups, however, were not significantly different:  $t(11.8) = -0.838$ ,  $p=0.419$ . Levene's test for this comparison was significant ( $p < 0.05$ ), so the *t*-value and *df* are reported for unequal variances. Thus, when participants with prior BBN training are removed from the analyses,

**Table 2** WiTalk palliative care workshop participants' experience with BBN

Parameter	Group A n (%)	Group B n (%)	Total n (%)	p-value
<b>Current year of training</b>				
Postgraduate Year 2	15 (100)	7 (58.3)	21 (80.8)	
Postgraduate Year 3		3 (25)	3 (11.5)	
Missing		2 (16.7)	2 (7.7)	
Total	15	12	27	
<b>BBN experience</b>	<b>Mean <math>\pm</math> SD</b>	<b>Mean <math>\pm</math> SD</b>	<b>Mean <math>\pm</math> SD</b>	
Your best estimate of the number of times in the past year that you were present when bad news was given to a personal family member, friend, or acquaintance	6.36 $\pm$ 12.6	6 $\pm$ 7.9	6.2 $\pm$ 10.7	0.2
Your best estimate of the number of times in the past year that you were present as a health professional when bad news was given to a patient	18.5 $\pm$ 13.6	15.1 $\pm$ 7.6	16.6 $\pm$ 10.4	0.8
Your best estimate of the number of times in the past year that you provided bad news to a patient	20.1 $\pm$ 24.8	8 $\pm$ 5.9	14.7 $\pm$ 19.4	0.4
Level of anxiety regarding breaking bad news to patients and families	3.14 $\pm$ 0.95	2.75 $\pm$ 0.62	2.96 $\pm$ 0.82	0.3

**Note:** Questions with mean (SD) and median (IQR); based on Mann–Whitney *U*-test for comparing the independent medians.

**Abbreviations:** BBN, breaking bad news; IQR, interquartile range.

**Table 3** Pre–post differences in MPathic-VR score

Group	Preseminar			Postseminar			t-test	df	p-value
	Min	Max	Mean $\pm$ SE	Min	Max	Mean $\pm$ SE			
Group A (n=15)	6.0	22.0	12.7 $\pm$ 1.2	4.0	14.0	8.1 $\pm$ 0.7	3.41 <sup>a</sup>	14	0.002
Group B (n=12)				4.0	18.0	10.3 $\pm$ 1.4			
Posttest only: Group A vs Group B							–1.40	16	0.175

**Notes:** Group A completed the VH simulation before and after the seminar; Group B completed the VH simulation only after the seminar. <sup>a</sup>Paired-samples *t*-test for preseminar and postseminar results for Group A only; <sup>b</sup>independent samples *t*-test for Group A and Group B posttest results only.

**Abbreviations:** Max, maximum; min, minimum; MPathic-VR, modeling professional attitudes and teaching humanistic communication in virtual reality.

**Table 4** Pre–post differences in MPathic-VR score for participants who reported no prior BBN training

	Preseminar			Postseminar			t-test	df	p-value
	Min	Max	Mean $\pm$ SE	Min	Max	Mean $\pm$ SE			
Group A (n=12)	7.0	22.0	14.0 $\pm$ 1.3	4.0	14.0	8.6 $\pm$ 0.8	3.32	11	0.007
Group B (n=9)				4.0	17.0	10.1 $\pm$ 1.6			
Posttest only: Group A vs Group B							–0.838	11.8	0.419

**Notes:** Group A completed the VH simulation before and after the seminar; Group B completed the VH simulation only after the seminar.

**Abbreviations:** BBN, breaking bad news; Max, maximum; min, minimum; MPathic-VR, modeling professional attitudes and teaching humanistic communication in virtual reality.

“both” sets of scores were significantly improved compared to the Group A preseminar test scores. Because the group without a pretest was not significantly different from the group with a pretest, the results support the hypothesis that the improvement was due to the seminar content rather than the pretesting. Moreover, it illustrates that MPathic-VR could be used to assess improvement.

The pre–post mean difference of 5.4 for participants without prior BBN training was slightly larger than the mean difference of 4.6 for participants with prior BBN. These results supported the final hypothesis that the difference was the greatest among participants without BBN training before participating in the seminar. The system scores reflected that hypothesis and detected a greater difference among those without prior BBN training.

## Discussion

The responsiveness of scores to the BBN training yields initial evidence of construct validity of MPathic-VR as an assessment tool for, in this instance, assessing performance-based competence in BBN among medical resident participants. The results offer support for the three hypotheses that MPathic-VR scores would improve from the pretest to the posttest state with exposure to training, that improvement would reflect the training rather than pretesting, and that the difference would be greatest among individuals without prior BBN training. Group A's postseminar MPathic-VR scores were comparable to Group B's postseminar test scores, and both sets of scores were significantly different from the preseminar test scores of Group A. Therefore, improved MPathic-VR scores after the seminar are a measure of



acquisition of knowledge and skills in delivering bad news to the VH patient.

Importantly, we found little evidence of pretest sensitization because the improvement of scores was not attributable to using MPathic-VR twice. The Group B posttest scores were lower than the Group A's pretest scores, and this difference reached statistical significance when participants with prior BBN experience were removed from the analyses. Moreover, the findings from our quasi-experimental two-group evaluation showed that scores on MPathic-VR could reliably distinguish between residents who did or did not have previous training in BBN. Presumably, individuals with previous BBN training would have also applied BBN communication techniques and gained experience. The MPathic-VR pre-seminar scores indeed demonstrated stronger competence for physicians with prior exposures to BBN training. These results suggest that MPathic-VR has utility for assessing BBN competence.

This study contributes to the understanding of VH-based assessment of communication skills by gathering validity evidence and examining the utility of VHs for assessment. First, the results demonstrate that MPathic-VR scores are indicators of BBN competence. Because assessment and learning are iterative and tightly interrelated, medical educators need consistent and reliable ways to formatively assess development of communication skills. MPathic-VR yields a reliable assessment, which can help to ensure appropriate instruction that is guided by the learner's demonstrated strengths or weaknesses. Second, this study supports MPathic-VR's value for standardized and efficient communication assessment. As introduced herein, SPIs provide a valid method of skill assessment but have substantial limitations when used for communication and professional assessment.<sup>14</sup> A computer-based method, such as MPathic-VR, can obviate those limitations by ensuring standardization of VH behaviors and learner-VH interaction. It can also ensure standardization over time, across multiple interactions, and across multiple institutions. Furthermore, it can avoid the logistical problems and expense that typify SPI programs. Thus, VH simulation is an important area of study to improve assessment of nontechnical skills and to provide training.

This study has potential limitations. One issue relates to the sample size. A post hoc power analysis for the current study demonstrated an actual power of  $\beta=0.956$  to detect the pre-post mean difference effect size ( $d=0.88$ ) that we found.<sup>45</sup> Nevertheless, future work needs to expand this research with larger samples and with other populations to

determine whether the effects observed here are generalizable, perhaps to fellow or attending physicians. To further build an even stronger validity argument, it would be informative to determine criterion relatedness by correlating the MPathic-VR scores with other measures of communication competence and to examine internal structures using factor analysis. Another limitation is that these results pertain to a particular assessment tool used for testing performance-based competence in BBN to a VH patient. Evidence of the validity of MPathic-VR for other assessment applications (eg, to assess interprofessional or intercultural communication) will require further validation studies.<sup>46</sup> The third limitation relates to the implementation of MPathic-VR as an assessment tool. Demonstrating the potential of interactive VH patients for competency assessment is an initial step, yet integrating it into a curriculum program will bring in implementation challenges that require further research. Assessments of other fields or skills may also be useful for assessing physician competence in cancer communication.

## Conclusion

Our findings provide initial construct validity evidence that the interactive VH program, MPathic-VR, can assess changes in BBN performance-based competence after completing BBN training. Further research is needed to gather evidence of criterion relatedness with other assessments and of construct validity with a larger sample. No single study can establish validity evidence for scores of an assessment, and multiple studies are needed to thoroughly develop a full validity argument.<sup>46</sup> To the best of our knowledge, these results appear to be the first to demonstrate that software featuring a VH patient interaction can be used for assessing communication skills in medical education.

## Abbreviations

BBN, breaking bad news; CBC, complete blood count; ER, emergency room; IRB, institutional review board; MPathic-VR, modeling professional attitudes and teaching humanistic communication in virtual reality; OSCEs, objective structured clinical examinations; SPIs, standardized patient instructors; VH, virtual human.

## Acknowledgments

Timothy C Guetterman, PhD, is an assistant professor in the Department of Family Medicine, University of Michigan. His research focus is enhancing health communication through the use of technology. His other focus is advancing the methodology of mixed methods research. Frederick W

Kron, MD, is a family medicine physician and president of Medical Cyberworlds, Inc. Dr Kron is an adjunct research investigator in the Department of Family Medicine, University of Michigan, MI, USA. Toby C Campbell, MD, MSCI, is an oncologist and associate professor in the Department of Medicine, School of Medicine and Public Health, University of Wisconsin–Madison, WI, USA. Mark W Scerbo, PhD, is a human factors psychologist and professor in the Department of Psychology, Old Dominion University, Norfolk, VA, USA. Amy B Zelenski, PhD, is Director of Education, Department of Medicine, University of Wisconsin–Madison, WI, USA. James Cleary, MD, is an oncologist and professor in the Department of Medicine, School of Medicine and Public Health, University of Wisconsin–Madison, WI, USA. Michael D Feters, MD, MPH, MA, is a family physician and professor in the Department of Family Medicine, University of Michigan, Ann Arbor, MI, USA. This work was supported by the National Institutes of Health, National Cancer Institute under grant number 1R43CA141987-01. The data sets analyzed in the current study are available from the corresponding author on reasonable request.

## Author contributions

All authors contributed toward data analysis, drafting and critically revising the paper and agree to be accountable for all aspects of the work.

## Disclosure

FWK serves as president and MDF has stock options in Medical Cyberworlds, Inc, the entity receiving Small Business Innovation Research I grant funds for this project. The University of Michigan Conflict of Interest Office considered the potential for conflict of interest and concluded that no formal management plan was required. The authors report no other conflicts of interest in this work.

## References

- Ong LM, De Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. *Soc Sci Med*. 1995;40(7):903–918.
- Hampton J, Harrison M, Mitchell J, Prichard J, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. *BMJ*. 1975;2(5969):486–489.
- Williams M, Hevelone N, Alban RF, et al. Measuring communication in the surgical ICU: better communication equals better care. *J Am Coll Surg*. 2010;210(1):17–22.
- Sutcliffe KM, Lewton E, Rosenthal MM. Communication failures: an insidious contributor to medical mishaps. *Acad Med*. 2004;79(2):186–194.
- Hulsman RL, Visser A. Seven challenges in communication training: learning from research. *Patient Educ Couns*. 2013;90(2):145–146.
- Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *AAMC. Acad Med*. 1993;68(6):443–451.
- Kneebone R, Bello F, Nestel D, Yadollahi F, Darzi A. Training and assessment of procedural skills in context using an integrated procedural performance instrument (IPPI). *Stud Health Technol Inform*. 2007;125:229–231.
- Neequaye SK, Aggarwal R, Van Herzele I, Darzi A, Cheshire NJ. Endovascular skills training and assessment. *J Vasc Surg*. 2007;46(5):1055–1064.
- Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. *Acad Med*. 1964;39(8):802–805.
- Reader T, Flin R, Lauche K, Cuthbertson BH. Non-technical skills in the intensive care unit. *Br J Anaesth*. 2006;96(5):551–559.
- Mitchell L, Flin R. Non-technical skills of the operating theatre scrub nurse: literature review. *J Adv Nurs*. 2008;63(1):15–24.
- Rall M, Gaba D. Patient simulators. In: Miller RF, editor. *Miller's Anesthesia*. 6th ed. New York: Elsevier; 2004:3073–3103.
- Morgan PJ, Kurrek MM, Bertram S, LeBlanc V, Przybyszewski T. Non-technical skills assessment after simulation-based continuing medical education. *Simul Healthc*. 2011;6(5):255–259.
- Turner TR, Scerbo MW, Gliva-McConvey GA, Wallace AM. Standardized patient encounters: periodic versus postencounter evaluation of nontechnical clinical performance. *Simul Healthc*. 2016;11(3):164–172.
- Newlin-Canzone ET, Scerbo MW, Gliva-McConvey G, Wallace AM. The cognitive demands of standardized patients: understanding limitations in attention and working memory with the decoding of nonverbal behavior during improvisations. *Simul Healthc*. 2013;8(4):207–214.
- Ekman P, Hager JC, Friesen WV. The symmetry of emotional and deliberate facial actions. *Psychophysiology*. 1981;18(2):101–106.
- Ekman P, Friesen WF. Felt, false and miserable smiles. *J Nonverbal Behav*. 1982;6(4):238–252.
- Laidlaw A, Salisbury H, Doherty EM, Wiskin C. National survey of clinical communication assessment in medical education in the United Kingdom (UK). *BMC Med Educ*. 2014;14(1):1–7.
- Cleland JA, Abe K, Rethans J-J. The use of simulated patients in medical education: AMEE Guide No. 42. *Med Teach*. 2009;31(6):477–486.
- Lok B. Teaching communication skills with virtual humans. *IEEE Comput Graph Appl*. 2006;26(3):10–13.
- Ferdig RE, Coutts J, DiPietro J, Lok B, Davis N. Innovative technologies for multicultural education needs. *Multicultur Educ Technol J*. 2011;1(1):47–63.
- Kron F. Medica ex machina: can you really teach medical professionalism with a machine? *J Clin Oncol*. 2007:692–697.
- Kenny P, Hartholt A, Gratch J, et al. Building interactive virtual humans for training environments. *Interservice/Industry Training, Simulation, and Education Conference (IITSEC)*. 2007. Available from: <http://ict.usc.edu/pubs/Building%20Interactive%20Virtual%20Humans%20for%20Training%20Environments.pdf>. Accessed June 21, 2017.
- Johnson WL, Friedland L. Integrating cross-cultural decision making skills into military training. In: Schmorrow D, Nicholson D, editors. *Advances in Cross-Cultural Decision Making*. London: Taylor & Francis; 2010:540–549.
- Lynch-Sauer J, VandenBosch T, Kron FW, et al. Nursing students' attitudes toward video games and related new media technologies: implications for nursing education. *J Nurs Educ*. 2011;50(9):513–523.
- Kron FW, Gjerde CL, Sen A, Feters MD. Medical student attitudes toward video games and related new media technologies in medical education. *BMC Med Educ*. 2010;10(1):50.
- Manton KG, Vaupel JW. Survival after the age of 80 in the United States, Sweden, France, England, and Japan. *N Engl J Med*. 1995;333(18):1232–1235.
- Johnson K, Stevens A, Lok B. The validity of a virtual human experience for interpersonal skills education. *SIGCHI Conference on Human Factors in Computing Systems*. San Jose, CA: 2007.

29. Andrade AD, Cifuentes P, Oliveira MC, Anam R, Roos BA, Ruiz JG. Avatar-mediated home safety assessments: piloting a virtual objective structured clinical examination station. *J Grad Med Educ*. 2011;3(4):541–545.
30. Tulskey JA. Beyond advance directives: importance of communication skills at the end of life. *JAMA*. 2005;294(3):359–365.
31. Orlander JD, Fincke BG, Hermanns D, Johnson GA. Medical residents' first clearly remembered experiences of giving bad news. *J Gen Intern Med*. 2002;17(11):825–831.
32. Ptacek JT, Ptacek JJ, Ellison NM. "I'm sorry to tell you." physicians' reports of breaking bad news. *J Behav Med*. 2001;24(2):205–217.
33. Salander P. Bad news from the patient's perspective: an analysis of the written narratives of newly diagnosed cancer patients. *Soc Sci Med*. 2002;55(5):721–732.
34. Parker PA, Baile WF, de Moor C, Lenzi R, Kudelka AP, Cohen L. Breaking bad news about cancer: patients' preferences for communication. *J Clin Oncol*. 2001;19(7):2049–2056.
35. Butow PN, Kazemi JN, Beeney LJ, Griffin AM, Dunn SM, Tattersall MH. When the diagnosis is cancer: patient communication experiences and preferences. *Cancer*. 1996;77(12):2630–2637.
36. Baile WF, Lenzi R, Parker PA, Buckman R, Cohen L. Oncologists' attitudes toward and practices in giving bad news: an exploratory study. *J Clin Oncol*. 2002;20(8):2189–2196.
37. Back AL, Arnold RM, Tulskey JA, Baile WF, Fryer-Edwards KA. Teaching communication skills to medical oncology fellows. *J Clin Oncol*. 2003;21(12):2433–2436.
38. Zhang B, Wright AA, Nilsson ME. Associations between advanced cancer patients' end-of-life conversations and cost experiences in the final week of life. *J Clin Oncol*. 2008;26(90150):9530.
39. Back AL, Arnold RM, Baile WF, et al. Efficacy of communication skills training for giving bad news and discussing transitions to palliative care. *Arch Intern Med*. 2007;167(5):453–460.
40. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES-A six-step protocol for delivering bad news: application to the patient with cancer. *Oncologist*. 2000;5(4):302–311.
41. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol*. 1995;50(9):741–749.
42. Back A. Communication between professions: doctors are from mars, social workers are from venus. *J Palliat Med*. 2000;3(2):221–222.
43. Back AL, Arnold RM. Discussing prognosis: "how much do you want to know?" talking to patients who are prepared for explicit information. *J Clin Oncol*. 2006;24(25):4209–4213.
44. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin; 2002.
45. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 3rd ed. New York: Academic Press; 1988.
46. Kane MT. An argument-based approach to validity. *Psychol Bull*. 1992;112(3):527–535.

## Advances in Medical Education and Practice

### Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

Dovepress

including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.