



Training and Validation of Incognito Standardized Patients for Assessing Oncology Fellows' Performance Regarding Breaking Bad News

Mandana Shirazi¹, Amir Hossein Emami ¹ and Afsaneh Yakhforoshha^{2,*}

¹Tehran University of Medical Sciences, Tehran, Iran

²Qazvin University of Medical Sciences, Qazvin, Iran

*Corresponding author: Department of Medical Education, Qazvin University of Medical Sciences, Qazvin, Iran. Tel: +98-283340153, Email: a.yakhforoshha@qums.ac.ir

Received 2021 January 23; Revised 2021 March 04; Accepted 2021 March 07.

Abstract

Background: Standardized patient (SP) has been applied to measure learner's communication challenges such as breaking bad news (BBN). When utilizing SP-based assessment, 2 steps should be considered in SP training; assessing SPs portrayal as the real patient (authenticity) and how SPs checklist fill out reproducibility.

Objectives: In this study, we described the process of training authentic and consistent SPs for evaluating oncology fellows' performance regarding BBN in Iran.

Methods: In this cross-sectional study, 8 eligible SPs took part in a 3-day educational meeting. Four different scenarios were developed regarding cancer patients along with corresponding checklists representing common presentations of illness. The accuracy of SPs portrayal was evaluated by experts, using a previously validated rating scale during observation of their role-playing. The reproducibility of SPs' portraits was measured, using a test-retest approach. The inter-rater agreement of the SPs' ability to fill out the BBN scale was measured by comparing the correlation between the SPs, who completed the scale, and oncologist faculty members' judgments, which is considered a gold standard.

Results: The findings of this study indicated that the cut-off score for the SPs' portrayal validity was 95%. The reliability of SPs portrayal was acceptable ($r = 0.89$). The inter-rater agreement between SPs and experts in filling the BBN scale ($k = 0.82$), as well as, the consistency of filling the BBN scale between SP groups were highly acceptable ($k = 0.86$).

Conclusions: The present study has demonstrated that if SP is trained appropriately, they have a high degree of reliability and validity to assess oncology fellows' performance regarding BBN skills.

Keywords: Standardized Patient, Breaking Bad News, Assessment

1. Background

Breaking bad news (BBN) is a complex communication task that is a common occurrence for clinicians working with oncology patients. Being aware of how bad news is shared between providers and patients is essential in the examination. Consequently, different formats of assessment tools (e.g. detailed checklist and global rating scales) have been developed, as well as the assessor (e.g. standardized patients and independent raters) (1).

Moreover, the need for valid assessment of medical students' preparation in communication skills has been emphasized in evidence. This issue is particularly important for the high-stakes exam, in which decisions of students' Pass/Fail are made (2). It was proposed that Standardized patients (SPs) play a vital role in ensuring more objective means for assessment, particularly in the field of CS (3).

Therefore, it is worthwhile to consider whether the use of SPs assessment has been also applied in North American medical schools and Licensure examinations (4, 5). At this point, certain studies have employed SPs for assessing CS and BBN (6-8).

Some studies indicated that adequately-trained SPs could be satisfactory alternatives to faculty raters (9, 10). On the contrary, other studies indicated less agreement between SPs' and instructors' ratings (11-13). Based on the previous research, intensive rater training is one of the most key points in achieving the consistency of scoring (14).

2. Objectives

A common concern regarding educators, who adopt the use of SPs for performance assessment, is to train SPs

reliably and consistently to re-enact scenarios and to assess the student. Despite the widespread use of SPs as an assessment tool, there is a paucity of literature in the oncology profession, seeking to determine the quality assurance, the accuracy (validity), and consistency (reliability) of the SPs as the tools used for the assessment of BBN. For this reason, this study assesses the psychometric characteristic of a well-trained and unannounced SP as the evaluator in the clinic.

3. Methods

This is a cross-sectional study to describe the process of creating valid SPs for assessing fellows' performance-related BBN. The study was conducted in the following steps:

3.1. Scenario Development

Four various clinical scenarios depicting cancer cases (chronic myeloid leukemia, chronic lymphocytic leukemia, lung cancer, and stomach cancer) were developed by consensus among the panel of experts, consisting of medical educationalists, and physicians specialized in oncology. Each scenario included the key points of standardized information related to BBN to be presented in the real encounter. To facilitate the scenario writing process, a case template was developed, in which the experts could fill out some detailed information regarding the case content consisting of social and demographic, lifestyle, symptoms, medical history, family history, findings on physical examination, laboratory and imaging. The content validity of the scenarios was ensured through consensus in an expert panel including 10 experts from the Medical Education and Department of Oncology at Tehran University of Medical Sciences (TUMS).

3.2. Selection of SPs

Eight SPs volunteers from the simulated patient pool at TUMS were invited to take part in this study.

3.3. Training the Standardized Patient

SPs received a 3-day training focusing on their realistic portrayal of each clinical scenario (authenticity), and SPs' practice rating the BBN's criteria on a checklist to record the practice of the performance of oncology correctly. The case developers and faculty members, who have experience in working with SPs, undertook the training of all SPs.

All SPs were trained by working as SP trainers.

Based on the SP training method (15), the following steps have been accomplished:

1) Familiarizing SPs with the clinical scenario reading the case materials;

- 2) Ensuring that SPs have learned trainer checklists;
- 3) Performing SPs role play of the cases with feedback from the trainers and peers;
- 4) Allowing SPs to dress rehearsal to enhance the authenticity of their performance;
- 5) Promoting SPs' ability of accurate and consistent checklist scoring skills.

During training (rating checklists and SPs' portrayals), the SPs read detailed written instructions describing the case, case-specific observational checklist, and watched videos regarding BBN, and also receive training protocols and complete assessment instruments.

Moreover, SPs play their roles with other SPs and faculty members under the supervision of an experienced faculty. In addition, the training process involved SPs-visit in the outpatient oncology setting. During these unannounced visits and interactions with real patients, the SPs became more familiar with all aspects of the scenario; besides, they learned how to handle certain situations. After the encounter in the oncology outpatient setting, SPs practiced all aspects of their roles as cancer patients.

Throughout the study, SPs participated in weekly meetings for further training in their roles, discussing common SP errors (rating checklist & portrayals), how to avoid them, and receiving feedback from both peers and faculty on their performances.

3.4. Validation of SPs' Portrayals

Based on different scenarios, case-specific and observational checklists for measuring SP role-plays were developed via an expert team consisting of oncologists and medical education experts. The validity of the scale was determined by agreement among the panel of experts. Each key relevant item that was noted to the SPs in training was included in the scale, comprising 5 items for verbal and 5 for nonverbal expressions. Each item was graded on a Likert scale from 1 to 3 (1 = poor, 2 = mild, and 3 = good" or "excellent). The final score was calculated as the mean score of the ten items.

One week after the training, the performance of each of the 8 SPs during the interactions with an oncologist as the doctor was assessed by other oncologists and a medical education expert, using case-specific scales. The SPs portrayed 2 cases with the same oncologist. Each of the 4 scenarios was performed by 2 SPs. The SPs' portrayal would have an acceptable accuracy if it reaches the cut-off score of 90% or above. To determine the reliability of SPs' portrayals, the test-retest approach across the first and second role-play was employed.

3.5. Validation of SPs' Completed the Checklist

BBN scale was selected as the tool for measuring the BBN skill of oncology fellows in the outpatient setting. The

BBN's validity and reliability were published in a previous study in Iran (16).

The BBN checklist had 16 items, measuring 7 variety domains of BBN skills, including the setting interview (3 items), strategy (2 items), planning (2 items), professionalism (1 item), empathy (2 items), knowledge (4 items), and invitation (2 items). This examination tool was to be completed by the SP after the consultation with the fellow.

For testing the concurrent validity of the SPs' ability to complete the checklists, each SP played their role with 3 oncologists (who were not involved in the study) as a clinician in a simulated medical environment. Afterward, SPs rated the clinician's performance. Following the SP-clinician encounter, other oncology faculty, who had experience in working as SP-based performance assessments independently, rated the SP-faculty interaction immediately. Accordingly, the agreement of the SP in grading the BBN scale and again scale-rating by an independent expert (oncologist faculty members) as a gold standard supported the concurrent validity of the SP-based performance assessment. Additionally, the inter-rater reliability of 2 SPs on the same condition was assessed for indicating the reliability of the SPs rating checklist.

3.6. Data Analysis

We analyzed the data using SPSS. To determine the correlation, the kappa (k) coefficient was computed.

3.7. Ethical Considerations

The study was approved by the Ethics Committee of Tehran University of Medical Sciences (Reference: IR.TUMS.REC.1394.1621).

4. Results

The mean age range of the SPs was 46.5 ± 9 . There were 2 men and 6 women, who had 9.5 ± 2.5 years of work experience. The mean score for measuring the SPs' portrayal validity was 2.97 (ranging: 2.90 - 3.00). The assessment of the portrayal's reliability indicated a mean of 89% (range: 82.6% - 96%) equal responses to the items on the observational rating scale. The mean K for the validity of SPs' filled-out checklists was 0.82 (ranging: 0.694 - 0.985). The mean K for the reliability of SPs' completed checklists was 0.86% (ranging: 0.725 - 1.000). The scores and individual test results for the 8 SPs are shown in [Table 1](#).

5. Discussion

Since outcome-based education and early clinical encounter is introduced in medical education, a great deal

has been accomplished in developing more standardized, objective, and structured methods of assessment. To address this need, SP-based assessment has become widespread in undergraduate and graduate medical education (17).

Although how SPs portray the case efficiently is identified as SP error in SP-based assessment, quality assurance is indicated to significantly reduce the SP performance error (18,19). In addition, literature has noted that SPs should be measured for accuracy and reliability of the portrayal ahead of being employed in performance assessment (20, 21).

In the present study, we evaluated the accuracy (validity) and consistency (reliability) of the SP-based performance assessment to estimate oncology fellows' performance regarding BBN skills, when SPs serve as unannounced patients in the clinical practice. According to evidence, the cut-off score of 85% or higher rate of agreement with the "expert" that developed the case for the SPs' portrayal is considered to have strong accuracy (22). Therefore, our results implied that SPs can present different scenarios of cancer patients with a high level of accuracy (cut-off score = 95% for all SPs). Our results were in line with the study representing 91% and 89% of accuracy for SP portrayal in the prenatal and cancer cases (23).

In our opinion, our finding is rather related to the development of a valid case-specific observational checklist to evaluate the SP's authenticity by the oncologist in defined task components. This also guided the oncologists to utilize such scale as an instructional tool to provide the SPs with constructive feedback.

On top of that, the current study indicated that the reliability of performance of all SPs portraying each scenario across two role-playings was acceptable. This might be due to the best selection of experienced SPs and systematic training of them, which is in concordance with the findings of an early study (24).

Regarding the validity and reliability of rating the fellows' performance on BBN checklists by SPs, in the current study, SPs' ratings were found to be highly consistent. Furthermore, there was a high percentage of agreement (80% -100%) between the SPs' completing the checklists and experts' judgments. It should be noted that the SP stake part in the training session learning how to rate BBN skill consistently. In addition, literature suggested that the reliability of the assessment by SPs increases using case-specific checklists and careful rater training (25). As noted in a previous study, providing rater training is of great value for achieving more reliable and valid results. However, one of the consequences of this experiment is that training not only ensures that all raters interpret the content of an item's description similarly, but also they apply similar standards to students' performances (13).

Table 1. Mean Scores of Validity of SPs' Portrayal and Inter-Rater Reliability of SPs' Portrayal as Well as the Validity and Reliability of Completed Checklists by Each SP

Standard Patient (SP)	Observational Rating Scale		Checklists	
	Total Mean of Raters' Validity of SPs' Portrayal	Percentage of Inter-rater Reliability of SPs' Portrayal	K Coefficient (validity of Completed Checklists)	K Coefficient (Reliability of Completed Checklists)
1st SP	3.000	93.8	0.848	0.921
2nd SP	2.96	96	0.794	1.000
3rd SP	2.98	79.8	0.784	0.725
4th SP	3.000	94	0.985	0.883
5th SP	3.000	90	0.784	0.788
6th SP	2.98	82.6	0.694	0.882
7th SP	2.90	83.	0.789	0.785
8th SP	2.95	92.8	0.890	0.898

Overall, our results demonstrate that well-trained incognito SP can serve as a standardized assessment tool for assessing oncology fellows' performances regarding BBN skills in the clinical field. Although the small sample size of our study limits the generalizability of the results, quality assurance processes for SP-based performance assessment of health care providers were outlined. It would be valuable for future research to report this process of training and validation of SPs in other areas of medicine.

Acknowledgments

The authors would like to thank all SPs, who have participated in this study.

Footnotes

Authors' Contribution: Study concept and design: Afsaneh Yakhforoshha. and Mandana Shirazi. Analysis and interpretation of data: Amir Hossein Emami. Drafting of the manuscript: Afsaneh Yakhforoshha and Mandana Shirazi.

Conflict of Interests: There are no conflicts of interest.

Ethical Approval: The study was approved by the Ethics Committee of Tehran University of Medical Sciences (reference: IR.TUMS.REC.1394.1621).

Funding/Support: There is no funding or support.

References

- Schildmann J, Kupfer S, Burchardi N, Vollmann J. Teaching and evaluating breaking bad news: a pre-post evaluation study of a teaching intervention for medical students and a comparative analysis of different measurement instruments and raters. *Patient Educ Couns*. 2012;**86**(2):210-9. doi: [10.1016/j.pec.2011.04.022](https://doi.org/10.1016/j.pec.2011.04.022). [PubMed: [21571487](https://pubmed.ncbi.nlm.nih.gov/21571487/)].
- Comert M, Zill JM, Christalle E, Dirmaier J, Harter M, Scholl I. Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE)-A Systematic Review of Rating Scales. *PLoS One*. 2016;**11**(3). e0152717. doi: [10.1371/journal.pone.0152717](https://doi.org/10.1371/journal.pone.0152717). [PubMed: [27031506](https://pubmed.ncbi.nlm.nih.gov/27031506/)]. [PubMed Central: [PMC4816391](https://pubmed.ncbi.nlm.nih.gov/PMC4816391/)].
- May W, Park JH, Lee JP. A ten-year review of the literature on the use of standardized patients in teaching and learning: 1996-2005. *Med Teach*. 2009;**31**(6):487-92. doi: [10.1080/01421590802530898](https://doi.org/10.1080/01421590802530898). [PubMed: [19811163](https://pubmed.ncbi.nlm.nih.gov/19811163/)].
- Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;**287**(2):226-35. doi: [10.1001/jama.287.2.226](https://doi.org/10.1001/jama.287.2.226). [PubMed: [11779266](https://pubmed.ncbi.nlm.nih.gov/11779266/)].
- Petrusa ER. Clinical Performance Assessments. *International Handbook of Research in Medical Education*. 2002. p. 673-709. doi: [10.1007/978-94-010-0462-6_26](https://doi.org/10.1007/978-94-010-0462-6_26).
- Ju M, Berman AT, Hwang WT, Lamarra D, Baffic C, Suneja G, et al. Assessing interpersonal and communication skills in radiation oncology residents: a pilot standardized patient program. *Int J Radiat Oncol Biol Phys*. 2014;**88**(5):1129-35. doi: [10.1016/j.ijrobp.2014.01.007](https://doi.org/10.1016/j.ijrobp.2014.01.007). [PubMed: [24661666](https://pubmed.ncbi.nlm.nih.gov/24661666/)].
- Lifchez SD, Redett R. A standardized patient model to teach and assess professionalism and communication skills: the effect of personality type on performance. *J Surg Educ*. 2014;**71**(3):297-301. doi: [10.1016/j.jsurg.2013.09.010](https://doi.org/10.1016/j.jsurg.2013.09.010). [PubMed: [24797843](https://pubmed.ncbi.nlm.nih.gov/24797843/)].
- Wang S, Shadrake L, Lyon MJ, Kim H, Yudkowsky R, Hernandez C. Standardized Patient-Based Assessment of Dermatology Resident Communication and Interpersonal Skills. *JAMA Dermatol*. 2015;**151**(3):340-2.
- Khan AS, Qureshi R, Acemoğlu H, Shabi-ul-Hassan S. Comparison of Assessment Scores of Candidates for Communication Skills in an OSCE, by Examiners, Candidates and Simulated Patients. *Creat Educ*. 2012;**3**(6):931-6. doi: [10.4236/ce.2012.326141](https://doi.org/10.4236/ce.2012.326141).
- Zanetti M, Keller L, Mazor K, Carlin M, Alper E, Hatem D, et al. Using standardized patients to assess professionalism: a generalizability study. *Teach Learn Med*. 2010;**22**(4):274-9. doi: [10.1080/10401334.2010.512542](https://doi.org/10.1080/10401334.2010.512542). [PubMed: [20936574](https://pubmed.ncbi.nlm.nih.gov/20936574/)].
- Panzarella KJ, Manyon AT. A model for integrated assessment of clinical competence. *J Allied Health*. 2007;**36**(3):157-64. [PubMed: [17941410](https://pubmed.ncbi.nlm.nih.gov/17941410/)].
- Rothman AI, Cusimano M. A comparison of physician examiners', standardized patients', and communication experts' ratings of international medical graduates' English proficiency. *Acad Med*. 2000;**75**(12):1206-11. doi: [10.1097/00001888-200012000-00018](https://doi.org/10.1097/00001888-200012000-00018). [PubMed: [11112723](https://pubmed.ncbi.nlm.nih.gov/11112723/)].
- Tasdelen Teker G, Odabasi O. Reliability of scores obtained from standardized patient and instructor assessments. *Eur J Dent Educ*. 2019;**23**(2):88-94. doi: [10.1111/eje.12406](https://doi.org/10.1111/eje.12406). [PubMed: [30450818](https://pubmed.ncbi.nlm.nih.gov/30450818/)].
- Chesser A, Cameron H, Evans P, Cleland J, Boursicot K, Mires G. Sources of variation in performance on a shared OSCE station across four UK medical schools. *Med Educ*. 2009;**43**(6):526-32. doi: [10.1111/j.1365-2923.2009.03370.x](https://doi.org/10.1111/j.1365-2923.2009.03370.x). [PubMed: [19493176](https://pubmed.ncbi.nlm.nih.gov/19493176/)].

15. Wallace P. *Coaching standardized patients: For use in the assessment of clinical competence*. Springer publishing company; 2007.
16. Farokhyar N, Shirazi M, Bahador H, Jahanshir A. Assessing the validity and reliability of spikes questionnaires regard in of medical residents awareness breaking bad news in TUMS 2012. *Razi J Med Sci*. 2014;**21**(122):29–36.
17. Gerzina HA, Stovsky E. *Standardized Patient Assessment Of Learners In Medical Simulation*. StatPearls. Treasure Island (FL); 2021. eng.
18. Baig LA, Beran TN, Vallevand A, Baig ZA, Monroy-Cuadros M. Accuracy of portrayal by standardized patients: results from four OSCE stations conducted for high stakes examinations. *BMC Med Educ*. 2014;**14**:97. doi: [10.1186/1472-6920-14-97](https://doi.org/10.1186/1472-6920-14-97). [PubMed: [24884744](https://pubmed.ncbi.nlm.nih.gov/24884744/)]. [PubMed Central: [PMC4035823](https://pubmed.ncbi.nlm.nih.gov/PMC4035823/)].
19. Sim AJ, Mahoney BP, Katz D, Reddy R, Goldberg A. The Clinical Educator Track for Medical Students and Residents. *The Comprehensive Textbook of Healthcare Simulation*. Springer; 2013. p. 575–85. doi: [10.1007/978-1-4614-5993-4_41](https://doi.org/10.1007/978-1-4614-5993-4_41).
20. Furman GE. The role of standardized patient and trainer training in quality assurance for a high-stakes clinical skills examination. *Kaohsiung J Med Sci*. 2008;**24**(12):651–5. doi: [10.1016/S1607-551X\(09\)70031-2](https://doi.org/10.1016/S1607-551X(09)70031-2). [PubMed: [19251561](https://pubmed.ncbi.nlm.nih.gov/19251561/)].
21. Levine AI, DeMaria S, Schwartz AD, Sim AJ. *The Comprehensive Textbook of Healthcare Simulation*. Springer Science & Business Media; 2013. doi: [10.1007/978-1-4614-5993-4](https://doi.org/10.1007/978-1-4614-5993-4).
22. Porcerelli JH, Brennan S, Carty J, Ziadni M, Markova T. Resident Ratings of Communication Skills Using the Kalamazoo Adapted Checklist. *J Grad Med Educ*. 2015;**7**(3):458–61. doi: [10.4300/JGME-D-14-00422.1](https://doi.org/10.4300/JGME-D-14-00422.1). [PubMed: [26457156](https://pubmed.ncbi.nlm.nih.gov/26457156/)]. [PubMed Central: [PMC4597961](https://pubmed.ncbi.nlm.nih.gov/PMC4597961/)].
23. Erby LA, Roter DL, Biesecker BB. Examination of standardized patient performance: accuracy and consistency of six standardized patients over time. *Patient Educ Couns*. 2011;**85**(2):194–200. doi: [10.1016/j.pec.2010.10.005](https://doi.org/10.1016/j.pec.2010.10.005). [PubMed: [21094590](https://pubmed.ncbi.nlm.nih.gov/21094590/)]. [PubMed Central: [PMC3158971](https://pubmed.ncbi.nlm.nih.gov/PMC3158971/)].
24. Shahidullah JD, Kettlewell PW. Using standardized patients for training and evaluating medical trainees in behavioral health. *Int J Health Sci Educ*. 2017;**4**(2):1–14.
25. Vu NV, Barrows HS. Use of Standardized Patients in Clinical Assessments: Recent Developments and Measurement Findings. *Educ Res*. 2016;**23**(3):23–30. doi: [10.3102/0013189x023003023](https://doi.org/10.3102/0013189x023003023).