Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: www.tandfonline.com/journals/imte20

Evaluating the performance of simulated patients: Development and validation of the Lisbon Assessment of Simulated Patients (LASP) scale

Paulo Oom, Sandra Oliveira, Leonor Bacelar-Nicolau, João Pereira, Rita Oom & Rodrigo Sousa

To cite this article: Paulo Oom, Sandra Oliveira, Leonor Bacelar-Nicolau, João Pereira, Rita Oom & Rodrigo Sousa (25 Mar 2026): Evaluating the performance of simulated patients: Development and validation of the Lisbon Assessment of Simulated Patients (LASP) scale, Medical Teacher, DOI: 10.1080/0142159X.2026.2641757

To link to this article: https://doi.org/10.1080/0142159X.2026.2641757

MEDICAL TEACHER
https://doi.org/10.1080/0142159X.2026.2641757

Evaluating the performance of simulated patients: Development and validation of the Lisbon Assessment of Simulated Patients (LASP) scale

Paulo Oom[a], Sandra Oliveira[b], Leonor Bacelar-Nicolau[a], João Pereira[a], Rita Oom[c] and Rodrigo Sousa[a]

[a]Center for Interdisciplinary Research in Health (CIIS), Católica Medical School, Universidade Católica Portuguesa, Lisboa, Portugal; [b]Católica Research Center for Psychological, Family, and Social Wellbeing, Faculty of Human Sciences, Universidade Católica Portuguesa, Lisboa, Portugal; [c]Católica Medical School, Universidade Católica Portuguesa, Lisboa, Portugal

ABSTRACT

Background: Effective communication between doctors and patients is crucial for positive health outcomes and patient satisfaction. In medical education, simulated patients (SPs) play a pivotal role in developing these skills. The quality of SPs' performances and the feedback they provide significantly influence the learning process. However, current assessment tools tend to be too lengthy, overlapping or focused solely on either role-play or feedback, failing to differentiate their individual impacts. This study aimed to develop and validate a concise, reliable instrument to assess the performance of SPs in both domains separately.

Methods: The Lisbon Assessment of Simulated Patients (LASP) scale was developed with expert input to ensure content validity. Data were gathered through 629 questionnaires completed by 180 medical students and 15 tutors, who assessed 25 SPs. To determine the factor structure and model fit, exploratory and confirmatory factor analyses were conducted on split subsamples. Internal consistency was assessed using Cronbach's alpha and McDonald's omega. Additionally, descriptive statistics and item-domain correlations were analysed.

Results: Factor analyses revealed a two-factor structure consisting of Role Play (4 items) and Feedback (6 items), which together accounted for 54% of the variance. The fit indices indicated an excellent model fit, with a CFI of 0.991, TLI of 0.988, and RMSEA of 0.020. The items demonstrated strong correlations with their respective domains, and the scale displayed good internal consistency, with a Cronbach's alpha of 0.82 overall, 0.72 for the Role Play domain, and 0.79 for the Feedback domain.

Discussion: The LASP scale is a practical and psychometrically robust tool designed to assess the performance of simulated patients. It evaluates both role-play and the quality of feedback, allowing for targeted improvements in simulated encounters and enhancing the effectiveness of communication training. This concise and validated instrument supports ongoing quality assurance and faculty development in medical simulation programs.

ARTICLE HISTORY
Received 6 January 2026
Accepted 3 March 2026

KEYWORDS
Simulated patients; medical education research; communication skills; role-play assessment; feedback quality

Introduction

Effective communication between doctors and patients is recognised as a fundamental skill in clinical practice, being associated with better health outcomes, increased patient satisfaction, and a reduction in medical errors [1–6].

In medical education, simulated patients (SP) play a critical role by providing students with opportunities to develop and refine communication skills in a safe and controlled environment [7,8]. However, the formative impact of such sessions greatly depends on the quality of the SP's role-play during the consultation, as well as the relevance and structure of the feedback provided to students after the simulation [9,10].

Despite their importance, few validated instruments systematically assess both SP role-play quality and the feedback process. Existing scales present several limitations: some are composed of an excessive number of items [11–16], making their use impractical in educational contexts; others contain overlapping or redundant items [11,15,16] or include items of questionable relevance [14,15]. Additionally, specific scales focus solely on either role-play or feedback [13,14,17], or provide a combined score that does not distinguish between the particular contributions of each domain to the overall outcome

CONTACT Paulo Oom poom@ucp.pt Center for Interdisciplinary Research in Health (CIIS), Católica Medical School, Universidade Católica Portuguesa, Lisboa, Portugal.

Practice points

- The LASP is a concise 10-item scale designed to assess the quality of role-play and feedback provided by simulated patients during medical student consultations.
- Factor analyses reveal a stable two-domain structure—Role Play and Feedback—accounting for 54% of the variance, with excellent model fit indices.
- The LASP demonstrates good internal consistency for both the total scale and its subscales, confirming its reliability as a measurement tool.
- The separate domain scores enable the targeted identification of strengths and weaknesses in simulated patient performance, thereby contributing to program improvement.
- This tool is practical for routine use and supports quality assurance and faculty development, highlighting the need for further validation in other settings.

[11,14–18]. In some cases, equal weight is assigned to role-play and feedback without considering their potentially distinct impact on the learning process [12,15,16].

This gap hinders the continuous monitoring and improvement of SP performance, ultimately limiting the potential of simulated encounters as educational strategies. The availability of a reliable and comprehensive scale would not only standardise and enhance the quality of SP performance but also promote more effective and constructive feedback for students, thereby supporting the acquisition of essential communication skills.

In this context, we have developed a new scale to evaluate SP performance in medical student consultations, specifically designed to assess both the quality of role-play during the encounter and the quality of feedback given at the end of the session. The scale comprises a reduced set of non-redundant items, each with independent scoring provided for its respective dimension.

This study aims to validate this new scale by analysing its psychometric properties and assessing its practical applicability in medical education settings.

Methods

The Board at Católica Medical School granted permission for the study to be conducted.

Participants

As part of our structured communication training program, we employ simulated patients (SPs) as an integral component of our educational strategy. Throughout the three-year bachelor phase of the medical curriculum, all students are systematically provided with opportunities to develop and consolidate their communication and physical examination skills through structured simulated patient encounters, during which the SP offers feedback to the student at the end of each consultation. Accordingly, the systematic evaluation of SP performance is essential to ensure and enhance the overall quality of the training program. To meet this need, we created a standardised instrument to evaluate the performance quality of individual SPs, based upon existing evaluation frameworks from other medical schools as reference models during the instrument's development process. A panel of experts was assembled to identify and validate the specific items of the scale, ensuring content validity and relevance.

At the beginning of the program, all students and simulated patients provided informed consent for their participation, were informed of the ethical principles regarding confidentiality and data collection and agreed to the use of the videotaped consultations for educational and investigation purposes.

After each SP consultation, students who were present during the consultation and their tutors, who watched the video recording of the consultation, complete the questionnaire to evaluate the SP's performance. Six hundred twenty-nine questionnaires, involving 25 SPs, 180 bachelor students, and 15 tutors, were analysed. Data was collected between September 2024 and March 2025. All ratings were collected anonymously, and no identifiable information about individual raters (including whether they were students or tutors) was linked to the study dataset. This approach aligns with the primary aim of the present study, which was to examine the internal structure and reliability of the LASP as an instrument for assessing simulated patient performance, rather than to compare stakeholder groups.

Instrument

The LASP: Lisbon Assessment of Simulated Patients (LASP) is an 11-item scale designed to evaluate individual simulated patients' (SPs') performance in role-playing and providing feedback. The questionnaire comprises two subscales measuring: 'Role Play' (four items), which reflects the SP's ability to perform realistically, and 'Feedback' (seven items), which refers to the SP's capacity to provide feedback in a neutral, concrete, and constructive way from the patient's perspective (Table 1). Items are rated on a five-point

Table 1. Initial items of the LASP instrument.

Items	Designation
Item 1	SP's appearance fits the role
Item 2	SP might be a real patient
Item 3	SP simulates physical complaints authentically
Item 4	SP is challenging/testing the student
Item 5	SP communicates how they felt during the consultation
Item 6	SP speaks about their experience of the consultation in the first person (uses I-statements)
Item 7*	SP's feedback is not related to medical content
Item 8	SP does not compare students with other students
Item 9	SP's feedback is neutral, and the student does not feel judged, neither in a positive nor in a negative way
Item 10	SP's feedback is concrete and constructive, giving examples from the consultation
Item 11	If allowed, SP provides extra feedback about issues that the student did not ask about

*Excluded after analysis.

Likert scale (from 1—Completely disagree to 5—Completely agree). High scores indicate higher levels of performance. Items 7 and 8 are reverse-coded. In addition to the items that make up the factorial structure of the measure, a global score for each domain ("Regarding the quality of the role portrayal, what mark would you give to the SP?"; "Regarding the quality of the feedback, what mark would you give to the SP?") may be used to capture the SP's overall perception of performance providing an integrative assessment of complementary value.

Data analysis

A descriptive statistical analysis was carried out to examine the item-response distribution (mean, SD, range, skewness and kurtosis), and box-plots were checked, as missing values were visually inspected. In a few cases where data were missing, a common feature of real-world datasets, values were imputed using the mean imputation method. Studies indicate that the limitations of mean imputation are minimal when less than 10% of the data is missing and when the correlations between variables are low [19,20].

To assess construct validity, an exploratory factor analysis (EFA) was performed, followed by a confirmatory factor analysis (CFA) and an analysis of item-domain correlations. As for the factor analytic procedures, the total sample was randomly divided into two subsamples. The first subsample ($n = 310$) was used to perform an exploratory factor analysis (EFA), with varimax rotation, according to the following criteria: (1) the Kaiser–Meyer–Olkin measure of sampling adequacy (KMO >0.60) and Bartlett's test of sphericity ($p < 0.05$); (2) the eigenvalue (>1) in line with the scree plot and the percentage of variance explained; and (3) communalities and factor loading of the variables (>0.50). To confirm the previous solution, CFA was conducted in the second subsample ($n = 319$). In CFA, each domain represented a latent variable, and each item was used as an observed variable. To analyse how well the model fits the sample data [21] the following recommended "goodness of fit" indices and respective cut-off values were used: the relative chi-square[$\chi^{2/}df \leq 2$ [22], Comparative Fit Index (CFI) $\geq$0.95; Goodness of Fit Index (GFI) $\geq$ 0.95, Tucker-Lewis Index (TLI) $\geq$ 0.95; Root Mean Square Error of Ap proximation(RMSEA) $\leq$0.05. Pearson's correlations were used to assess item and domain correlations. The ratio of subjects to items was high (1:31), indicating an adequate number of subjects for each variable and, therefore, more accurate solutions [23].

To assess the internal consistency of the scale and the unidimensionality of each factor, in addition to Cronbach's alpha, the McDonald's omega test was performed, as it is considered the most accurate estimate of reliability [24].

The statistical analyses were conducted in SPSS, version 30, except for the EFA and CFA, which were performed using JASP, version 0.16.1.

Results

Descriptive analyses

The means and standard deviations (SDs) for the two subscales and the total scale are as follows: Role Play: 4.71 ($SD = 0.43$); Process of Feedback: 4.72 ($SD = 0.40$); Total Scale: 4.72 (SD = 0.35). The means for the individual items ranged from 4.41 to 4.88, with standard deviations ranging from 0.43 to 1.18.

Construct validity

The EFA results indicated adequate values for the KMO measure of adequacy (KMO = 0.846) and Bartlett's test of sphericity ($p < .001$), confirming that the sample was suitable for data processing. The factor structure yielded a two-factor solution, accounting for 54% of the variance, as confirmed by the scree test.

Regarding the item loadings, the first factor, 'Role Play', comprised four items, explaining 28.7% of the variance. The second factor, 'Feedback', comprised six items explaining 25.3% of the variance. Due to low factor loading (<0.40), item 7, 'SP's feedback is not related to medical content', was excluded from the new analysis. The best factor structure consisted of two domains and 10 items.

The CFA results confirmed the factor structure identified by the EFA analysis.

According to the recommended cut-off values for appropriate data-model fit [25], the relative chi-square fit index ($\chi^2/df = 1.13$) was considered adequate.

Acceptable values were also found for the 'goodness of fit' indices: ($\chi^2 = 37.3$; $p > 0.05$).

Examining the modification indices improved the model fit, allowing the residual covariances between items 5 and 9 to be correlated. The results revealed excellent model fit indices: CFI = 0.991, TLI = 0.988, RMSEA = 0.020 and SRMR = 0.037, which suggests a relatively good fit between the hypothesised model and the observed data [26].

Pearson's correlations, displayed in Table 2, showed that each item correlated positively and significantly (p <.01) within its own domain, with no item correlating more strongly with another domain than with its own. The item-domain correlation coefficients were high, ranging from 0.749 to 0.755 for the Role Play domain and from 0.569 to 0.765 for the Feedback domain.

Reliability

Both reliability indices indicated that the LASP scale was adequate. The Cronbach's alpha (α) and McDonald's omega (ω_t) values for the total scale were: 0.82 and 0.83, respectively. The Cronbach's alpha (α) and McDonald's omega (ω) values for the Role Play and Feedback subscales were 0.72 and 0.76, and 0.79 and 0.80, respectively, demonstrating the unidimensionality of each subscale.

The LASP instrument

The final instrument consisted of two domains (Role-play and Feedback) and 10 items. For each domain, a global score was asked (Table 3).

Table 2. Correlations of items/domains of the LASP scale.

Items and domains	Role play	Feedback
Item 1	**0.749****	0.448**
Item 2	**0.755****	0.382**
Item 3	**0.751****	0.376**
Item 4	**0.753****	0.247**
Item 5	0.395**	**0.712****
Item 6	0.352**	**0.761****
Item 8	0.140*	**0.569****
Item 9	0.315**	**0.696****
Item 10	0.443**	**0.714****
Item 11	0.318**	**0.765****

Correlation coefficients of item/domain are shown in bold.
*$p < 0.05$.
**$p < 0.01$.

Discussion

This study aimed to validate a new scale for evaluating simulated patient (SP) performance by examining its psychometric properties. The evaluation was based on data collected from a large sample of feedback from both students and tutors. The primary objective was to create a practical tool that provides a comprehensive assessment of SP performance during medical student consultations.

The LASP demonstrates robust validity and reliability according to all recommended psychometric criteria. Exploratory and confirmatory factor analyses supported a clear two-factor solution, comprising Role Play (4 items) and Feedback (6 items). Item 7 was removed from the final version due to low factor loading and conceptual inconsistency, as it asks whether SP feedback is related to medical content, which reflects substantive accuracy rather than feedback quality. Fit indices exceeded conventional thresholds (CFI = 0.991, TLI = 0.988, RMSEA = 0.020), indicating an excellent match between the hypothesised two-domain model and the empirical data [26]. The scale's composite reliability was also high, with Cronbach's alpha values of 0.82 for the overall scale, 0.72 for Role Play, and 0.79 for Feedback (McDonald's omega values confirm these results), demonstrating both internal consistency and domain-specific coherence.

While the results of the present study add to a growing body of research findings, supporting the use of the LASP scale to systematically assess the quality of SP performance in medical education, there are some limitations to be addressed. First, mean imputation was used to handle missing data on the scale items. Although the level of missing data is relatively low (8.1%), mean imputation may reduce variability and underestimate standard errors. Future studies should address this methodological shortcoming. Second, the nested data structure may inflate model fit indices in CFA. In the present analysis, strict anonymity prevented the inclusion of rater-level identifiers, precluding multilevel modelling. Future research should account for clustering using multilevel approaches. A larger number of SPs and a more diverse range of raters are also

Table 3. The LASP instrument.

Role play	1	SP's appearance fits the role
	2	SP might be a real patient
	3	SP simulates physical complaints authentically
	4	SP is challenging/testing the student
	score	Regarding the quality of the role portrayal, what mark would you give to the SP?
Feedback	1	SP communicates how they felt during the consultation
	2	SP speaks about their experience of the consultation in the first person (uses I-statements)
	3	SP does not compare students with other students
	4	SP's feedback is neutral, and the student does not feel judged, neither in a positive nor in a negative way
	5	SP's feedback is concrete and constructive, giving examples from the consultation
	6	If allowed, SP provides extra feedback about issues that the student did not ask about
	score	Regarding the quality of the feedback, what mark would you give to the SP?

recommended to strengthen the construct validity and generalisability of the LASP measure. Further research should also assess measurement invariance between students and tutors to control for potential structural differences between these groups and to ensure that the scale operates equivalently across raters. Third, the high mean item scores may indicate a potential ceiling effect, possibly due to the high level of SP training represented in this study sample. Future studies with larger and more diverse samples in terms of SPs' experience are crucial to strengthen the generalisability and robustness of the scale. In addition, employing longitudinal designs would allow for the assessment of temporal stability and test–retest reliability of the LASP scale. Despite these methodological limitations, this study provides evidence that the LASP scale is structurally valid and internally consistent, supporting its suitability for research and medical education.

The new instrument not only evaluates core aspects of the simulated medical encounter but also distinguishes itself by systematically assessing both the quality of role-play (realism, challenge, authenticity) and the feedback process (constructiveness, neutrality, and specificity of feedback) after the simulation. Although some previously published scales have also addressed both domains [11,12,15,16,18], others have focused exclusively on either role-play [13,17] or feedback [14], failing to capture the integrated nature of SP performance in communication training. Compared with most established instruments, the LASP offers distinct advantages. Whereas the MaSP [18] uses 20 items and the NESP [11] 28 items and both deliver aggregate scores that combine role-play and feedback quality, the LASP provides a comprehensive assessment with only 10 items, yielding separate domain scores for Role Play and Feedback. This separation facilitates precise diagnosis of specific SP strengths and weaknesses, enabling targeted faculty development rather than undifferentiated retraining, and directly addresses a key limitation of prior scales.

Correlation analysis revealed that each item aligned most strongly with its own domain and did not exhibit excessive redundancy or cross-loading, indicating the non-overlapping structure of the scale. Additionally, no item was found to be redundant or irrelevant, as indicated by the strong item-domain correlations (ranging from 0.749 to 0.755 for Role Play and from 0.569 to 0.765 for Feedback). The final instrument comprises just 10 discriminative items, a number significantly lower than most scales currently in use [11–16]. This allows the LASP to be efficiently administered without burdening raters, whilst providing clearly interpretable results. In our view, this balance between practical feasibility and psychometric robustness is crucial for instruments intended for routine use, as overly long checklists, even if comprehensive, are less likely to be completed consistently in real-world clinical teaching environments. Importantly, unlike many other instruments [11,14–18], LASP calculates scores for each dimension independently, rather than through an aggregate approach. This means individual strengths and areas for improvement in both domains—Role Play and Feedback—can be distinctly recognised.

Notably, and in contrast to other instruments that evaluate both domains [11,12,15,16], the feedback dimension encompasses a greater number of discriminative items than the role-play dimension. While authentic role-play enhances realism and can increase student engagement and motivation [27], it is the quality of the feedback that enables students to identify gaps, reflect on their communication skills, and make corrections, ultimately leading to sustained learning improvements [28]. These design choices position the LASP not merely as an evaluative checklist but as a tool that operationalises key principles of dialogic and learner-centred feedback by supporting feedback conversations grounded in the SP's lived experience of the encounter. In doing so, the instrument may also contribute to strengthening students' feedback literacy by repeatedly exposing them to feedback that they must actively interpret and integrate into their ongoing development of communication skills. While highlighting the central pedagogical importance of SP feedback in learner development, this approach enables a more detailed assessment of feedback quality, facilitating the pinpointing of specific areas that require improvement.

Ultimately, the LASP provides a practical and comprehensive solution, responding to a recognised need for integrated, concise and domain-specific assessment of SPs. Its concise item set, strong psychometric foundation, and separation of domains provide robust support for program quality improvement and faculty development. Regular LASP monitoring thus supports ongoing quality improvement, enabling programme directors to allocate training resources efficiently and to track the effectiveness of interventions over time.

Local cultural factors, curriculum design, education context and specific norms in the training of simulated patients may influence the results of this study. Cultural differences in feedback preferences, such as a greater emphasis on indirectness, relational face-saving, or hierarchical deference in some contexts, may affect the relevance, rating patterns, or even the factor structure of items. These contextual factors limit the generalisability of the findings, emphasising the importance of replication across different sites and the need for multicentric research.

In this study, the LASP was developed and validated within a communication-focused undergraduate curriculum. Although designed for this context, aspects such as the emphasis on role-play authenticity and the quality of feedback from the simulated patient's perspective could also apply to other simulation types, including interprofessional or procedural scenarios. However, in these settings, the LASP would need further validation to reflect team interactions and the balance between technical and non-technical skills.

In the broader context of medical education, SPs remain essential for enabling students to learn and practice communication skills safely and effectively. SP performance, when continuously, systematically and closely monitored and guided by concrete, reliable scores, such as LASP, facilitates the establishment and maintenance of high standards in both the realism of clinical encounters and the effectiveness of feedback [9,10]. In turn, this contributes to more constructive learning environments and supports the professional growth of both SP and medical students [7,8].

Author contributions

CRediT: **Paulo Oom**: Conceptualization, Data curation, Investigation, Project administration, Writing – original draft, Writing – review & editing; **Sandra Oliveira**: Conceptualization, Formal analysis, Writing – review & editing; **Leonor Bacelar-Nicolau**: Conceptualization, Formal analysis, Writing – review & editing; **João Pereira**: Conceptualization, Writing – review & editing; **Rita Oom**: Conceptualization, Investigation, Writing – review & editing; **Rodrigo Sousa**: Conceptualization, Investigation, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partially supported by FCT - Fundação para a Ciência e Tecnologia, I.P. by project reference UID/04279/2025 and DOI identifier https://doi.org/10.54499/UID/04279/2025 - Centro de Investigação Interdisciplinar em Saúde.

Notes on contributors

Paulo Oom, Center for interdisciplinary Research in Health (CIIS), Católica Medical School, Universidade Católica Portuguesa. Lisboa. Portugal.

Sandra Oliveira, Católica Research Center for Psychological, Family, and Social Wellbeing, Faculty of Human Sciences, Universidade Católica Portuguesa. Lisboa. Portugal.

Leonor Bacelar-Nicolau, Center for interdisciplinary Research in Health (CIIS), Católica Medical School, Universidade Católica Portuguesa. Lisboa. Portugal.

João Pereira, Center for interdisciplinary Research in Health (CIIS), Católica Medical School, Universidade Católica Portuguesa. Lisboa. Portugal.

Rita Oom, Católica Medical School, Universidade Católica Portuguesa. Lisboa. Portugal.

Rodrigo Sousa, Center for interdisciplinary Research in Health (CIIS), Católica Medical School, Universidade Católica Portuguesa. Lisboa. Portugal.

ORCID

Paulo Oom http://orcid.org/0000-0002-5329-8782
Sandra Oliveira http://orcid.org/0000-0003-1378-9744
Leonor Bacelar-Nicolau http://orcid.org/0000-0003-0421-1262
João Pereira http://orcid.org/0000-0002-8800-9784
Rodrigo Sousa http://orcid.org/0000-0001-8239-2666

Data availability statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Samir MM, Eisa NM, Khaled H, et al. The impact of communication skills training on patient-centered attitude and patient satisfaction: a longitudinal study of Egyptian physicians. Prev Med Rep. 2025;55:103117. doi: 10.1016/j.pmedr.2025.103117
2. Singhal S, Shah RB, Bansal S, et al. Doctor–patient communication practices: a cross-sectional survey on Indian physicians. J Family Med Prim Care. 2024;13(11):5198–5206. doi: 10.4103/jfmpc.jfmpc_945_24
3. Chen X, Liu C, Yan P, et al. The impact of doctor-patient communication on patient satisfaction in outpatient settings: implications for medical training and practice. BMC Med Educ. 2025;25(1):830. doi: 10.1186/s12909-025-07433-y
4. Leotin S. Addressing health communication gaps: improving patient experiences and outcomes through human-centered design. J Patient Exp. 2025;12:23743735251334015. doi: 10.1177/23743735251334015
5. Danaher TS, Berry LL, Howard C, et al. Improving how clinicians communicate with patients: an integrative review and framework. J Serv Res. 2023;26(4):493–510. doi: 10.1177/10946705231190018
6. Sharkiya SH. Quality communication can improve patient-centred health outcomes among older patients: a rapid review. BMC Health Serv Res. 2023;23(1):886. doi: 10.1186/s12913-023-09869-8
7. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE guide no 42. Med Teach. 2009;31(6):477–486. doi: 10.1080/01421590903002821
8. Flanagan OL, Cummings KM. Standardized patients in medical education: a review of the literature.

Cureus. 2023;15(7): E 42027. doi: 10.7759/cureus.42027
9. Lewis KL, Bohnert CA, Gammon WL, et al. The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). Adv Simul. 2017;2(1):10. doi: 10.1186/s41077-017-0043-4
10. Oom P. The contribution of simulated patients to undergraduate medical education: a pathway to educational excellence. Acta Med Port. 2025;38(8):484–490. doi: 10.20344/amp.22902
11. Bouter S, Van Weel-Baumgarten E, Bolhuis S. Construction and validation of the nijmegen evaluation of the simulated patient (NESP): assessing simulated patients' ability to role-play and provide feedback to students. Acad Med. 2013;88(2):253–259. doi: 10.1097/ACM.0b013e31827c0856
12. Perera J, Perera J, Abdullah J, et al. Training simulated patients: evaluation of a training approach using self-assessment and peer/tutor feedback to improve performance. BMC Med Educ. 2009;9(1):37. doi: 10.1186/1472-6920-9-37
13. Resende KA, Cavaco AM, Luna-Leite MA, et al. Training and standardization of simulated patients for multicentre studies in clinical pharmacy education. Pharm Pract (Granada). 2020;18(4):2038. doi: 10.18549/PharmPract.2020.4.2038
14. Schlegel C, Woermann U, Rethans JJ, et al. Validity evidence and reliability of a simulated patient feedback instrument. BMC Med Educ. 2012;12(1):6. doi: 10.1186/1472-6920-12-6
15. Wind LA, Van Dalen J, Muijtjens AMM, et al. Assessing simulated patients in an educational setting: the MaSP (Maastricht Assessment of Simulated Patients). Med Educ. 2004;38(1):39–44. doi: 10.1111/j.1365-2923.2004.01686.x
16. Himmelbauer M, Seitz T, Seidman C, et al. Standardized patients in psychiatry – the best way to learn clinical skills? BMC Med Educ. 2018;18(1):72. doi: 10.1186/s12909-018-1184-4
17. Pritchard SA, Keating JL, Nestel D, et al. Physiotherapy students can be educated to portray realistic patient roles in simulation: a pragmatic observational study. BMC Med Educ. 2020;20(1):471. doi: 10.1186/s12909-020-02382-0
18. Gonullu I, Dogan CD, Erden S, et al. A study on the standard setting, validity, and reliability of a standardized patient performance rating scale – student version. Ann Med. 2023;55(1):490–501. doi: 10.1080/07853890.2023.2168744
19. Joel LO, Doorsamy W, Paul BS. A comparative study of imputation techniques for missing values in healthcare diagnostic datasets. Int J Data Sci Anal. 2025;20(7):6357–6373. doi: 10.1007/s41060-025-00825-9
20. Tsikriktsis N. A review of techniques for treating missing data in OM survey research. J Oper Manage. 2005;24(1):53–62. doi: 10.1016/j.jom.2005.03.001
21. Hooper D, Coughlan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. Electron J Bus Res Methods. 2008;6(1):53–60.
22. Schreiber JB, Nora A, Stage FK, et al. Reporting structural equation modeling and confirmatory factor analysis results: a review. J Educ Res. 2006;99(6):323–338. doi: 10.3200/JOER.99.6.323-338
23. Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval. 2005;10(1):1–9. doi: 10.7275/jyj1-4868
24. Irwing P, Hughes DJ. Test development. In: Irwing P, Booth T, Hughes DJ, editors. The Wiley handbook of psychometric testing. New Jersey: John Wiley & Sons; 2018. p. 1–47. doi: 10.1002/9781118489772.ch1
25. Ullman JB. Structural equation modelling. In: Tabachnick B, Fidell L, editors. Using multivariate statistics. 7th ed. Always learning. New York: Pearson; 2019. p. 528–612.
26. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equat Model: Multidiscipl J. 1999;6(1):1–55. doi: 10.1080/10705519909540118
27. Gorski S, Prokop-Dorner A, Pers M, et al. The use of simulated patients is more effective than student role playing in fostering patient-centred attitudes during communication skills training: a mixed method study. Biomed Res Int. 2022;2022(1):1498692. doi: 10.1155/2022/1498692
28. Elendu C, Amaechi DC, Okatta AU, et al. The impact of simulation-based training in medical education: a review. Medicine (Baltimore). 2024;103(27): E 38813. doi: 10.1097/MD.0000000000038813