Journal of Inonu University Health Services Vocational School
İnönü Üniversitesi Sağlık Hizmetleri Meslek Yüksekokulu Dergisi

e-ISSN: 2147-7892 Volume 13, Issue 3 (2025) 766-777 doi: 10.33715/inonusaglik.1657842

Research Article

EVALUATION OF BASIC INTERVIEW SKILLS REALIZED WITH SIMULATED PATIENT APPLICATIONS

Giray KOLCU[1] *Mukadder İnci BAŞER KOLCU*[2]
[1,2]*Süleyman Demirel University, Faculty of Medicine, Isparta*

Article Info

Received: *14 March 2025*
Accepted: *13 August 2025*

Keywords

Communication skills,
Interview skills,
Medical education,
Simulated patient.

ABSTRACT

Effective communication skills are crucial in healthcare and directly affect patient outcomes and satisfaction. This study aimed to evaluate the interview skills training and scoring process integrated into the training programs of the Süleyman Demirel University Faculty of Medicine Term 3 students. A cross-sectional descriptive quantitative research design was employed, utilizing an interview skills assessment scale developed by simulated patient trainers. Scoring data from student interviews conducted in the 2022-23 and 2023-24 academic years were collected and analyzed. The results revealed significant differences in central tendency and distribution characteristics among the assessment groups, highlighting potential inconsistencies in the scoring models. Peer and self-assessment scores exhibited left-skewed and leptokurtic distributions, which suggest that most scores were higher with fewer lower scores. Simulated patient evaluations showed varying patterns, with SP1 and SP4 scores indicating stricter but consistent scoring, whereas SP2 scores demonstrated a wider range with a tendency toward lower values. The low Cohen's kappa values across the assessment groups suggested inconsistencies in evaluation practices. The scale's reliability, assessed using McDonald's ω and Cronbach's α, indicated a high internal consistency. The study showed that there were differences between raters despite their training aimed at standardizing evaluation criteria. These findings highlight the need for improved rater training, standardization of assessment criteria, and regular calibration sessions to increase the fairness and reliability of performance evaluations. Future research should investigate the factors contributing to assessment differences and aim to calibrate the scoring criteria across different assessment methods.

INTRODUCTION

Effective communication skills play a fundamental role in healthcare and directly affect patient outcomes, satisfaction, and overall quality of care (Mercan, Özcan & Aydın, 2018b; Sarıkaya, Uzuner, Gülpınar, Keklik & Kalaça, 2004). The development and evaluation of these skills in medical education are of great importance for the training of competent health professionals (Mercan, Özcan & Aydın, 2018a). One of the most effective methods for teaching and evaluating communication skills is the simulated patient (SP) approach (Mercan et al., 2018a; Sarıkaya et al., 2004; Terzi, 2001). SP-based education provides students with a controlled and structured learning environment, giving them the opportunity to experience

Mukadder İnci BAŞER KOLCU ✉, mukadderkolcu@sdu.edu.tr
Süleyman Demirel University, Faculty of Medicine, Isparta

clinical interactions, receive feedback, and develop their skills before they encounter real patients (Bearman & Nestel, 2014).

SP practice is widely used in medical and health sciences education to help students develop clinical reasoning, professionalism, and patient-centered communication skills (Mercan et al., 2018a). These practices were based on realistic medical encounters with students by trained individuals in specific scenarios. Through standardized interactions and assessment criteria, SP-based training ensures that communication skills are consistently measured by different students.

Assessment of communication skills is vital for improving patient interactions in healthcare and for effectively training healthcare professionals. In this context, the systematic measurement of areas such as verbal and non-verbal communication, active listening, empathy, and professionalism stand out as important tools that improve patient outcomes as well as educational interventions. Structured rating scales used to improve the communication skills of healthcare professionals allow for reliable and valid assessment of these skills (Camerlynck et al., 2022; Carrasco-Guirao et al., 2024; Mahmoud et al., 2023).

Current literature shows that simulation training increases participants' communication skills and confidence in patient care (Camerlynck et al., 2022; Carrasco-Guirao et al., 2024). Zolnierek and Dimatteo emphasized the importance of communication in the doctor-patient relationship to increase patient compliance (Zolnierek & DiMatteo, 2009). In addition, Mahmoud et al. highlighted the positive effects of communication on the physician-patient relationship and collaboration within healthcare teams (Mahmoud et al., 2023). This shows the effectiveness of educational strategies in improving the communication skills of health professionals (Carrasco-Guirao et al., 2024). Many studies have emphasized the importance of SP training in the assessment and development of communication skills (Kruijver et al., 2001; Walsh, 2015; Yıldırım Sarı & Doğan, 2022).

The SP Laboratory, established at Süleyman Demirel University in 2019, provides an important infrastructure for the development of communication skills in medical education (Kolcu & Başer Kolcu, 2024). The laboratory consists of three examination rooms and one control room and is staffed with plural trained and experienced SPs. This laboratory was subjected to a self-assessment process and evaluated within the scope of Association for Simulated Practice in Healthcare (ASPIH) standards, and it was determined that it largely complied with these standards (Diaz-Navarro C, Laws-Chapman C, Moneypenny M & Purva M, 2023; Kolcu & Başer Kolcu, 2025). In our laboratory, interviewing skills training, bad news training, and various thesis studies have been conducted (Kolcu & Başer Kolcu, 2024).

Within the scope of the pre-graduation medical education program at Süleyman Demirel University Faculty of Medicine, interview skills training was provided to third-year students before their transition to the clinic (Keiser & Turkelson, 2017). The instructional process was structured in three stages: pre-briefing, SP interviews, and debriefing. During the first stage, students were informed about basic interview skills and shown a training video. Students then transformed their theoretical knowledge into practical applications by conducting one-to-one interviews with plural SPs. After the interviews, they were graded by their peers, themselves, and by SP (Bartlett, Pace, Arora & Penm, 2023; Howley, 2004; Özan & Yurdabakan, 2008). After scoring, structured feedback was provided and a debriefing process was conducted. This structured training model helps students develop effective communication skills and prepare them for clinical practice.

This study aimed to evaluate the interview skills training and scoring process integrated into the training programs of the Süleyman Demirel University Faculty of Medicine third-year students. In addition, by examining the reliability and validity of the scale, we have contributed to the development of methods to measure the communication skills of health professionals in an objective and standardized manner. These findings offer significant implications for optimizing SP-based training programs and elevating the overall quality of medical education.

MATERIAL AND METHOD

This study used a cross-sectional descriptive quantitative research design. Considering the ethical dimensions of the study, approval number 76/3 was obtained from the relevant ethics committee on 29.05.2024. The interview skills assessment scale developed by SP trainers was used in this study. This scale was designed to assess the interview skills of health professionals and consisted of 20 items. Each item was scored on four levels: 0 (not observed), 1 (not performed), 2 (performed at the novice level), and 3 (performed at the master level). The total score was calculated as a minimum of 20 and a maximum of 60 after the evaluation of all items.

Our SP education program includes several issues related to scoring practices. In addition, students were given preliminary information about scoring themselves and their peers before scoring within the scope of the study.

Within the scope of this research, scoring data from student interviews conducted in the 2022-2023 and 2023-2024 academic years were collected. These data were analyzed using Jamovi and JASP software. In the statistical analyses, descriptive statistics were calculated for the scale items. The consistency of the items in the evaluation process was determined by examining their mean, standard deviation, and distribution characteristics. Cohen's kappa was

also calculated to evaluate the difference between raters. The Cronbach's alpha coefficient was calculated to assess the reliability of the scale. The findings provide information about the validity and reliability of the measurement tool used in the evaluation of interview skills. To this end, we investigated whether SP practices constitute an objective and reliable evaluation process.

RESULT

In this study, the means, standard deviations, skewness, and kurtosis values of 613 scores belonging to the different evaluation groups were analyzed (N = 613). The mean score in the peer scoring group (n=128) was 56.61±4.60. The skewness value was -1.97, indicating that the distribution was skewed to the left. The kurtosis value was 4.48, indicating that the distribution was more pointed (leptokurtic) than normal. In the self-scoring group (n=38), the mean was 55.68±4.53. The skewness value was -1.79, indicating a left-skewed distribution. The kurtosis value was 3.84, indicating that the distribution was more pointed than normal. In the SP1 group (n=90), the mean score was 55.62±5.97. The skewness value was 2.59, indicating significant left skewness. The kurtosis value was 9.57, indicating a highly pointed distribution. In the SP2 group (n=130), the mean score was 42.42±10.27. The skewness value is 0.66, indicating a right-skewed distribution. The kurtosis value was -1.05, indicating that the distribution was flattened (platykurtic). In the SP3 group (n=101), the mean score was 53.93±4.48. The skewness value of -0.43 indicated an almost symmetrical distribution. The kurtosis value was -0.56, indicating that the distribution was slightly flatter than normal. In the SP4 group (n=126), the mean score was 54.69±7.32. The skewness value was -2.84, indicating a highly left-skewed distribution. The kurtosis value is 8.87, indicating a highly pointed distribution. In general, the distributions of peer scoring, self-scoring, SP1, and SP4 groups are left-skewed and pointed, while the SP2 group shows a right-skewed and flat distribution. The distribution of the SP3 group was closer to the normal distribution than those of the other groups. The Shapiro-Wilk test results showed $p<0.001$ in all groups, indicating that the distributions were not normal. This may necessitate the use of nonparametric methods in the analysis (Table 1).

Table 1. Groups Mean Scores and Distributions

					Skewness		Kurtosis		Shapiro-Wilk	
	N	Mean	±	SD	Skewness	SE	Kurtosis	SE	W	p
Peer score	128	56.61	±	4.60	-1.97	0.21	4.48	0.42	0.75	< .001
Self-score	38	55.68	±	4.53	-1.79	0.38	3.84	0.75	0.82	< .001
SP1	90	55.62	±	5.97	-2.59	0.25	9.57	0.50	0.72	< .001
SP2	130	42.42	±	10.27	0.66	0.21	-1.05	0.42	0.84	< .001
SP3	101	53.93	±	4.48	-0.43	0.24	-0.56	0.48	0.95	< .001

SP4	126	54.69	±	7.32	-2.84	0.22	8.87	0.43	0.66	< .001
Total score	613	52.56	±	8.70	-1.35	0.10	0.83	0.20	0.80	< .001

When the mean scores and standard deviations of the scale items evaluating interviewing skills were examined, it was observed that the participants generally received high scores. Greeting (2.78±0.46), showing interest and respect (2.80±0.43), and initiating the interview (2.78±0.46) received the highest scores. This finding suggests that the participants successfully applied their basic communication skills in patient interviews. The items of introducing oneself and one's role (2.45±0.77), clarifying (2.27±0.80) and summarizing (2.25±0.80) had the lowest mean scores. This finding indicates that advanced communication skills such as clarifying and summarizing were performed less effectively during the interview process. Critical communication skills, such as language use skill (2.73±0.48) and effective listening (2.75±0.46), received high scores. This finding indicates that the participants were successful in using clear language and actively listening to the patient. The high scores when the patient is included in the process (2.69±0.52) and when an empathic approach is shown (2.69±0.53) indicate that the participants attach importance to patient-centered communication. However, it can be said that the patient should be more involved in the process and the empathic approach should be further strengthened. The total score of 52.56±8.70 indicates that the overall interview skills are high, but there is room for improvement in some areas. It can be said that especially clarification and summarization skills need to be strengthened (Table 2).

Table 2. Interview Skill Assessment Scale Scores

Items	**Mean**	±	**Sd**
1. Greeting	2.78	±	0.46
2. Introducing oneself and explaining the role	2.45	±	0.77
3. Body language and eye contact, appropriate tone	2.73	±	0.51
4. Demonstrating interest and respect	2.80	±	0.43
5. Ensuring the patient's physical comfort	2.71	±	0.52
6. Initiating the conversation	2.78	±	0.46
7. Collecting socio-demographic information	2.50	±	0.67
8. Investigating the patient's chief complaint and medical history	2.74	±	0.52
9. Inquiring about past medical history	2.65	±	0.61
10. Gathering family history	2.58	±	0.67
11. Asking open-ended questions	2.54	±	0.71
12. Language proficiency (speed, fluency, clarity)	2.73	±	0.48
13. Active listening (avoiding interruptions, appropriate body language, maintaining eye contact; offering condolences or well wishes if needed)	2.75	±	0.46
14. Clarification (rephrasing the same statement with different words)	2.27	±	0.80
15. Summarizing	2.25	±	0.80
16. Engaging the patient in the process	2.69	±	0.52
17. Demonstrating empathy	2.69	±	0.53
18. Explaining the next steps of the procedure or treatment	2.74	±	0.51
19. Closing the conversation	2.53	±	0.71
20. Time management	2.69	±	0.56
Total score	52.56	±	8.70

The mean scores and standard deviations of the different assessment groups for the specific items are shown. In the peer and self-scoring groups, the means generally ranged between 2.60 and 2.90, with no major differences between the two groups. This suggests that there is significant consistency between participants' self-ratings and their peers' ratings. In the SP1 and SP3 groups, the average values were generally greater than 2.50. However, in some items, the evaluations of these groups remained at lower levels than those of the other groups. The overall averages of the SP4 group were distributed similarly to those of these groups. The SP2 group differed significantly from the other groups in that their scores were generally lower. It had the lowest averages, especially in terms of clarifying, summarizing, and ending the interview. This suggests that the SP2 group had difficulties with these skills. In general, the items with the highest scores were related to basic communication skills. Items such as greeting, showing interest and respect, and effective listening were rated highly by all the groups. However, the low average scores of the SP2 group indicate that this group, in particular, may need to improve their communication skills. The peer and self-assessment results were quite similar, indicating a generally consistent assessment process (Table 3).

Table 3. Comparison of Interviewing Skills Assessment Scale Items and Raters

	Peer scoring		**Self-scoring**		**SP1**		**SP2**		**SP3**		**SP4**	
	Mean	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**	**Mean**	**Sd**
1	2,91	0.29	2,92	0.27	2,88	0.39	2,32	0.52	2,96	0.20	2,87	0.50
2	2,63	0.67	2,66	0.71	2,68	0.47	2,11	0.71	2,36	0.89	2,45	0.86
3	2,87	0.34	2,82	0.46	2,82	0.41	2,24	0.57	2,91	0.32	2,86	0.49
4	2,96	0.19	2,95	0.23	2,93	0.25	2,32	0.50	2,96	0.20	2,86	0.47
5	2,88	0.35	2,89	0.31	2,78	0.44	2,22	0.56	2,88	0.38	2,79	0.51
6	2,92	0.27	2,89	0.31	2,92	0.31	2,29	0.53	2,96	0.20	2,87	0.46
7	2,70	0.51	2,71	0.57	2,53	0.66	2,06	0.68	2,40	0.71	2,74	0.55
8	2,91	0.28	2,82	0.39	2,89	0.35	2,18	0.64	2,87	0.34	2,88	0.45
9	2,80	0.45	2,79	0.47	2,90	0.37	2,15	0.65	2,62	0.65	2,79	0.55
10	2,75	0.56	2,76	0.59	2,83	0.48	2,11	0.67	2,57	0.65	2,65	0.70
11	2,84	0.43	2,74	0.50	2,71	0.52	1,70	0.78	2,75	0.43	2,75	0.55
12	2,84	0.37	2,79	0.47	2,88	0.36	2,25	0.48	2,90	0.30	2,85	0.47
13	2,91	0.29	2,84	0.37	2,89	0.35	2,28	0.50	2,87	0.34	2,83	0.45
14	2,75	0.50	2,66	0.53	2,60	0.56	1,65	0.81	2,00	0.88	2,27	0.69
15	2,73	0.51	2,58	0.64	2,52	0.60	1,73	0.81	1,89	0.88	2,30	0.68
16	2,80	0.42	2,79	0.41	2,67	0.58	2,26	0.52	2,87	0.37	2,84	0.46
17	2,87	0.34	2,82	0.39	2,63	0.53	2,19	0.56	2,90	0.33	2,85	0.47
18	2,89	0.34	2,82	0.39	2,90	0.34	2,24	0.55	2,88	0.35	2,83	0.52
19	2,79	0.46	2,61	0.64	2,87	0.40	1,96	0.77	2,49	0.80	2,62	0.64
20	2,87	0.36	2,84	0.44	2,79	0.49	2,15	0.62	2,88	0.38	2,79	0.53

It shows Cohen's kappa values and their confidence intervals calculated to measure the agreement between different evaluation groups. In general, Cohen's kappa values are low, indicating weak agreement between assessors. In particular, the negative Cohen's kappa value

between peer assessment and self-assessment (-0.028) indicated that there was no significant agreement between the two assessment methods. It is also noteworthy that the agreement between SP groups was low. For example, the Cohen's kappa values between SP1 and SP2 (0.009) and SP1 and SP3 (-0.009) suggest that the agreement is almost random. The highest agreement was observed between peer assessment and SP4 (0.213); however, this value indicated only weak to moderate agreement (Table 4). The results suggest inconsistencies between assessors and low overlap between the different assessment methods. This may indicate areas for improvement, such as assessor training or standardization of the assessment criteria.

Table 4. Rater Agreement

Cohen's kappa

			95% CI	
Ratings	Cohen's kappa	SE	Lower	Upper
Average Cohen's kappa	0.034			
Peer-scoring - Self scoring	-0.028	0.028	-0.082	0.026
Peer-scoring - SP1	0.104	0.040	0.026	0.182
Self-scoring - SP1	0.016	0.044	-0.070	0.101
Peer-scoring - SP2	0.026	0.017	-0.008	0.060
Self-scoring - SP2	-6.930×10^{-4}	9.546×10^{-4}	-0.003	0.001
SP1 - SP2	0.009	0.010	-0.010	0.029
Peer-scoring - SP3	0.115	0.036	0.044	0.186
Self-scoring - SP3	-0.034	0.010	-0.054	-0.014
SP1 - SP3	-0.009	0.024	-0.057	0.038
SP2 - SP3	-0.008	0.002	-0.013	-0.004
Peer-scoring - SP4	0.213	0.037	0.141	0.285
Self-scoring - SP4	0.009	0.035	-0.059	0.077
SP1 - SP4	-5.231×10^{-4}	0.023	-0.046	0.045
SP2 - SP4	0.043	0.021	0.002	0.083
SP3 - SP4	0.054	0.033	-0.010	0.118

The reliability of the scale evaluated in this study was examined using McDonald's ω and Cronbach's α. Cronbach's α and McDonald's ω values were 0.954 and 0.951, respectively. Both measures indicated that the scale had high internal consistency. The 95% confidence intervals calculated to assess the stability of the reliability estimates were also narrow. While the lower limit for Cronbach's α was 0.948 and the upper limit was 0.959, the range for McDonald's ω was 0.946–0.957 (Table 5).

Table 5. Reliability Analyses

Estimate	McDonald's ω	Cronbach's α
Point estimate	0.951	0.954
95% CI lower bound	0.946	0.948
95% CI upper bound	0.957	0.959

DISCUSSION

This study examined the distribution of scores in different assessment groups, including peer, self-, and SP assessments, which are important areas of debate in the current SP practice literature (Bokken, Linssen, Scherpbier, van der Vleuten & Rethans, 2009; Güllüdere, Yardım, Sezik & Şenol, 2014; Ljungman & Silén, 2008; Özan & Yurdabakan, 2008; Taylor, Haywood & Shulruf, 2019; Şenol & Başarıcı, 2014).

The results of our study reveal significant differences in central tendency and dispersion characteristics between the groups, in line with the current literature, and highlight potential inconsistencies in scoring models (Kolcu & Başer Kolcu, 2023; Taylor et al., 2019). The findings of the study show that although many topics related to scoring exist in the training programs for SP and students are provided with preliminary information about scoring themselves and their peers before the assessments, differences between some raters may still occur. In the literature, the importance of organizing training for scoring standardization to reduce the difference between scorers is mentioned (Barrett, 2001; Hubert, 1977; Koo & Li, 2016; McHugh, 2012; Rae, 1988; Warrens, 2010). However, as seen in our study, differences may occur even in SPs who have received long-term education.

Peer assessment scores exhibited a mean of 56. 61±4. 60, with a skewness of -1. 97 and kurtosis of 4.48. These values indicated a left-skewed and leptokurtic distribution, suggesting that most scores were concentrated at the higher end of the scale. Similarly, self-assessment scores (55.68±4.53) demonstrated a left-skewed distribution (-1.79) with a high kurtosis value (3.84), indicating a peaked distribution. These findings suggest that both peer and self-assessments tend to yield higher scores, potentially reflecting leniency in scoring or an overestimation of performance. Among the SP evaluations, SP1 scores (55.62 ± 5.97) were notably left-skewed (-2.59) with a high kurtosis value (9.57), indicating a concentration of higher scores with a sharp peak. SP4 scores (54.69±7.32) followed a similar pattern, with a skewness of -2.84 and kurtosis of 8.87. These findings suggest that these groups might have a stricter but more consistent scoring pattern, resulting in uniform and elevated scores. In contrast,

SP2 scores (42.42±10.27) showed a right-skewed distribution (0.66) with a platykurtic kurtosis value (-1.05), suggesting a wider range of scores with a tendency toward lower values. SP3 scores (53.93±4.48) exhibited a nearly normal distribution with minimal skewness (-0.43) and a slightly platykurtic nature (-0.56), making this group the closest to a normal distribution.

The low Cohen's kappa values across various assessment groups suggest significant inconsistencies in evaluation practices. The negative agreement between peer and self-assessments indicates potential issues with reflective practice among trainees as they may not accurately assess their abilities or performance in alignment with their peers. Similarly, the random agreements noted among the SP groups point to potential flaws in the assessment methodologies or training.

The reliability results reveal that the items of the scale largely measure the same construct, and the assessments are reliable.

These findings underscore the necessity of ensuring consistency in assessment criteria and training evaluators to reduce scoring variability. The significant differences observed between the peer/self-assessment and SP evaluations may reflect variations in perceived competency, evaluation standards, or scoring tendencies. The relatively normal distribution of the SP3 scores suggests that this group may represent a more balanced assessment approach.

This study had several limitations that should be considered. First, the sample sizes of the assessment groups were unequal, which may have influenced the statistical power of the comparisons. Second, the study relied on data from a single institution, which limits the generalizability of the findings to other educational settings. Third, the potential influence of subjective biases on peer and self-assessments could not be fully controlled. Finally, while skewness and kurtosis values provide insights into distribution characteristics, they do not capture all nuances of assessment variability. Future studies should aim to replicate these findings with larger and more diverse samples and explore additional factors that may contribute to assessment discrepancies.

Future research should explore the factors contributing to these variations, such as assessor training, rubric clarity, and influence of prior experience on scoring patterns. Additionally, efforts should be made to calibrate the scoring criteria across different evaluation methods to enhance fairness and reliability in performance assessments.

CONCLUSION

These findings suggest a need for improvement in areas such as assessor training, standardization of assessment criteria, and regular calibration sessions. Improving training for

both peer assessors and SP assessors can increase consistency by providing a better understanding of assessment criteria. Establishing clear and standardized criteria can help reduce ambiguity, leading to a more holistic approach to assessment and minimizing scoring discrepancies. Implementing regular calibration sessions for assessors can increase the level of agreement between different types of assessments by ensuring consistency in the assessment approaches.

Acknowledgements

We thank Süleyman Demirel University for supporting this research. The university's facilities and resources have played a crucial role in this regard. We are grateful to our faculty members, research assistants, and the technical staff for their assistance. We believe that our study makes a significant contribution to the field of medical education.

Author Contributions: Concept - GK, MİBK; Design – MİBK; Supervision - GK; Resources- MİBK; Materials - MİBK; Data Collection and/or Processing - GK; Analysis and/or Interpretation - GK, MİBK; Literature Review - GK, MİBK; Writing - GK, MİBK; Critical Review - GK, MİBK

Conflict of Interest: The authors declare no conflicts of interest.

Financial Disclosure: The authors declare that this study received no financial support.

REFERENCES

Barrett, S. (2001). The impact of training on rater variability. *International Education Journal*, *2*(1), 49–58.

Bartlett, A., Pace, J., Arora, A., & Penm, J. (2023). Self-Reflection and Peer-Assessments Effect on Pharmacy Students’ Performance at Simulated Counselling Sessions. *Pharmacy*, Vol. 11. https://doi.org/10.3390/pharmacy11010005

Bearman, M., & Nestel, D. (2014). *Simulated Patient Methodology: Theory, Evidence and Practice*. https://doi.org/10.1002/9781118760673.ch5

Bokken, L., Linssen, T., Scherpbier, A., van der Vleuten, C., & Rethans, J.-J. (2009). Feedback by simulated patients in undergraduate medical education: a systematic review of the literature. *Medical Education, 43*(3), 202–210. https://doi.org/10.1111/j.1365-2923.2008.03268.x

Camerlynck, M., Saunders, A., Attoe, C., Williams, L. M., Cross, S., & Wilson, C. (2022). Interprofessional Simulation Training for Perinatal Mental Health: A Mixed Methods Study. *Reproductive Female and Child Health*, *1*(2), 122–131. https://doi.org/10.1002/rfc2.17

Carrasco-Guirao, J. J., Leal-Costa, C., Castaño-Molina, M. de los Á., Ferrer, M. B. C., Molina-Rodríguez, A., Agea, J. L. D., & Martínez, M. (2024). Exploring How Evidence-Based Practice, Communication, and Clinical Simulation Outcomes Interact in Nursing Education: A Cross-Sectional Study. *Nursing Reports*, 14(1), 616–626. https://doi.org/10.3390/nursrep14010047

Diaz-Navarro C, Laws-Chapman C, Moneypenny M, & Purva M. (2023). *The ASPiH Standards - 2023: guiding simulation-based practice in health and care*.

Güllüdere, H. H., Yardım, S., Sezik, M., & Şenol, Y. (2014). Akran yardımı ile eğitimin tıp eğitiminde kullanımı. *Tıp Eğitimi Dünyası*, 13(39), 19–25.

Howley, L. D. (2004). Performance assessment in medical education: Where we've been and where we're going. *Evaluation and the Health Professions*, 27(3), 285–303. https://doi.org/10.1177/0163278704267044

Hubert, L. (1977). Kappa revisited. *Psychological Bulletin*, 84(2), 289.

Keiser, M. M., & Turkelson, C. (2017). Using students as standardized patients: Development, implementation, and evaluation of a standardized patient training program. *Clinical Simulation in Nursing*, 13(7), 321–330.

Kolcu, G., & Başer Kolcu, M. (2024). Bad News Training in SDU.

Kolcu, G., & Başer Kolcu, M. İ. (2023). Comparison of Different Raters Scores in Simulated Patient Training Program. *Internatıonal Journal of Interdıscıplınary Interactıon In Health Scıences*, 2(2), 45–54.

Kolcu, G., & Başer Kolcu, M. İ. (2024). *Süleyman Demirel Üniversitesi Tıp Fakültesi Simüle Hasta Laboratuvarı*. Akademisyen Kitabevi.

Kolcu, G., & Başer Kolcu, M. İ. (2025). Evaluation of Süleyman Demirel University Faculty of Medicine Simulated Patient Applications within the Scope of The ASPiH Standards 2023. *Medical Journal of Süleyman Demirel University*, 32. https://doi.org/10.17343/sdutfd.1506549

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kruijver, I. P. M., Kerkstra, A., Kerssens, J. J., HoItkamp, C. C. M., Bensing, J. M., & Wiel, H. B. M. van de. (2001). Communication Between Nurses and Simulated Patients With Cancer: Evaluation of a Communication Training Programme. *European Journal of Oncology Nursing*, *5*(3), 140–150. https://doi.org/10.1054/ejon.2001.0139

Ljungman, A. G., & Silén, C. (2008). Examination involving students as peer examiners. *Assessment & Evaluation in Higher Education*, *33*(3), 289–300. https://doi.org/10.1080/02602930701293306

Mahmoud, M. B., Badawey, J. T., Kouki, A. M. K., Malik, L. F., Shakally, S. M. S., & Nasser, A. A. (2023). *Effect of Educational Campaign on Interpersonal Communication Skills Among Health Care Professionals: Pre-Post Study, Jeddah, KSA*. https://doi.org/10.21203/rs.3.rs-3553849/v1

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282.

Mercan, N., Özcan, C. T., & Aydın, M. S. (2018a). Psikiyatride ve İletişim Eğitiminde Simüle Hasta Uygulamaları. *Current Approaches in Psychiatry / Psikiyatride Guncel Yaklasimlar*, 10(3), 292–301. https://doi.org/10.18863/pgy.364112

Mercan, N., Özcan, C. T., & Aydın, M. S. (2018b). Psikiyatride ve İletişim Eğitiminde Simüle Hasta Uygulamaları. *Psikiyatride Guncel Yaklasimlar - Current Approaches in Psychiatry*, 10(3), 292–301. https://doi.org/10.18863/pgy.364112

Özan, S., & Yurdabakan, İ. (2008). Öz ve Akran Değerlendirmenin Temel İletişim Becerileri Başarısı Üzerindeki Etkileri. *Tıp Eğitimi Dünyası*, 27(27), 27–39.

Rae, G. (1988). The Equivalence of Multiple Rater Kappa Statistics and Intraclass Correlation Coefficients. *Educational and Psychological Measurement*, 48(2), 367–374. https://doi.org/10.1177/0013164488482009

Sarıkaya, Ö., Uzuner, A., Gülpınar, M. A., Keklik, D., & Kalaça, S. (2004). İletişim becerileri eğitimi: İçerik ve değerlendirme. *Tıp Eğitimi Dünyası*, 14(14).

Şenol, Y., Başarıcı, S. (2014). Student Opinions About Standardized Patient Practices: First-Year Results. *Journal*, 13(41), 19–26.

Taylor, S., Haywood, M., & Shulruf, B. (2019). Comparison of effect between simulated patient clinical skill training and student role play on objective structured clinical examination performance outcomes for medical students in Australia. *Journal of Educational Evaluation for Health Professions*, 16, 3. https://doi.org/10.3352/jeehp.2019.16.3

Terzi, C. (2001). *Toplum sağlığına bir köprü: Tıp eğitimi.* İstanbul: İletişim Yayınları.

Walsh, K. (2015). The Future of Simulation in Medical Education. *Journal of Biomedical Research*. https://doi.org/10.7555/jbr.29.20140138

Warrens, M. J. (2010). Inequalities between multi-rater kappas. *Advances in Data Analysis and Classification*, 4, 271–286.

Yıldırım Sarı, H., & Doğan, P. (2022). Öğrencilerin Görme Engelli Simüle Hasta ile İletişim Becerilerinin Değerlendirilmesi: Pilot Çalışma. *Journal of Infant, Child and Adolescent Health*, 2(1), 1–10.

Zolnierek, K. B. H., & DiMatteo, M. R. (2009). Physician Communication and Patient Adherence to Treatment. *Medical Care*, 47(8), 826–834. https://doi.org/10.1097/mlr.0b013e31819a5acc