



Validation of Checklists and Evaluation of Clinical Skills in Cases of Abdominal Pain With Simulation in Formative, Objective, Structured Clinical Examination With Audiovisual Content in Third-Year Medical Students' Surgical Clerkship

JJ Ruiz-Manzanera, PhD,[†] J Almela-Baeza, PhD,[‡] A Aliaga, MD,[†] G Ádanez, PhD,[§] F Alconchel, PhD,[†] JM Rodríguez, PhD,[†] F Sánchez-Bueno[†], P Ramírez[†] and, B Febrero, PhD[†]

[†]General Surgery Service, Virgen de la Arrixaca Hospital. Murcia, Spain, Department of Surgery, Pediatrics, Obstetrics and Gynecology, University of Murcia, Instituto Murciano de Investigación Biosanitaria (IMIB) Pascual Parrill, El Palmar, Murcia, Spain; [‡]Department of Communication, University of Murcia, Espinardo, Murcia, Spain; and [§]Department of Medicine, University of Murcia, Espinardo, Murcia, Spain

OBJECTIVES: The objective of this study was to develop and validate 6 checklists for evaluating abdominal pain in clinical simulation scenarios; to assess student competencies in managing 6 clinical cases using OSCE, based on faculty evaluations; and to analyze discrepancies between faculty and student evaluations.

DESIGN: A practical workshop was designed to address 6 clinical scenarios of abdominal surgical conditions. Four scenarios employed medium fidelity simulators, while 2 scenarios employed standardized patient methodology. Prior to the workshop, students received theoretical audiovisual material. At the conclusion of the workshop, students were evaluated using checklists that assessed communication, privacy, anamnesis, and technical skills. Ten workshops were conducted over 3 years, using the OSCE (Objective Structured Clinical Examination) format for evaluation.

SETTING: In the statistical analysis, t-Student tests or ANOVA were employed to ascertain whether there were any significant differences between the groups. In the process of validating checklists for clinical scenarios, 6 experts were asked to evaluate each item on a scale of 1

to 9. To assess the degree of agreement among experts, the intraclass correlation coefficient (ICC) was employed.

PARTICIPANTS: The study involved a total of 670 third-year medical students from the University of Murcia (UMU), Spain, who participated in the subject "Medical-Surgical Skills."

RESULTS: High levels of appropriateness were observed for the checklist items, with mean scores above 7.5 points, as well as high levels of inter-expert agreement. Students obtained a mean score of 8 points in the evaluation of each clinical scenario. No significant differences were found between faculty and student scores ($p < 0.05$).

CONCLUSIONS: The learning method focused on clinical scenarios of abdominal surgical diseases effectively enhanced the clinical skills of third-year medical students. It used pre-existing audiovisual materials, hands-on workshops with medium-fidelity simulators, and standardized patients. Consistent evaluations from students and faculty confirmed the efficacy of these strategies. (J Surg Ed 81:1756–1763. © 2024 The Author(s). Published by Elsevier Inc. on behalf of Association of Program Directors in Surgery. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>)

Correspondence: Javier Almela-Baeza, PhD, Department of Communication, Faculty of Communication and Documentation of University of Murcia, C. Campus Universitario, 8, 30100 Murcia, Spain; e-mail: javier.almela@um.es

KEYWORDS: Clinical simulation, audiovisual contents, medical students, abdominal pain case, check list validation

INTRODUCTION

The systematic approach to abdominal pain is a fundamental component of the training of undergraduate medical students in their initial years of clinical training.¹⁻⁵ This learning should extend beyond the acquisition of skills related to abdominal physical examination and encompass essential competencies such as communication with the patient, anamnesis, and the development of skills in diagnostic and therapeutic management.^{2,6-9} This approach contributes to the development of clinical reasoning skills as early as the first years of the medical degree.

Furthermore, it is crucial to create diverse scenarios that enable students to expand their knowledge and skills in relation to the most prevalent diagnoses, facilitating an appropriate differential diagnosis.^{8,10-15} In this context, the use of medium-fidelity abdominal simulators offers a broader range of possibilities for the development of clinical scenarios, in comparison to those performed with a standardized patient. The simulator enables the simulation of pathological aspects of abdominal physical examination, including inspection, auscultation, percussion, and palpation.¹⁶⁻¹⁹ This approach facilitates the strengthening of skills in the management of patients presenting with abdominal pain.^{1,2,9,12}

The implementation of the Objective Structured Clinical Examination (OSCE) represents a pivotal educational instrument within the undergraduate medical training process.^{8,20} This methodology allows for the objective assessment of the diverse competencies in specific clinical scenarios.²¹⁻²⁴ A substantial body of research has demonstrated that formative OSCEs with immediate feedback have a significant impact on students' self-confidence and clinical competence in specific areas, creating a less intimidating and safer learning environment.^{10,20-24,16}

The utilization of validated checklists that encompass the requisite competencies is imperative for the assessment conducted via OSCE.^{6,10} However, it is equally important to consider the assessment performed by different examiners, as previous studies have demonstrated that discrepancies in the assessment of different items may exist between different assessors.^{6,21} Moreover, it is crucial to consider the perspective of other students. Peer feedback and evaluation can be beneficial, as they offer a distinct perspective and can contribute to the development of communication skills, teamwork, and self-reflection in students.^{25,26}

This study had the following objectives: 1) Development and validation of 6 checklists for the evaluation of abdominal pain in different clinical simulation scenarios; 2) Evaluation of student competencies in the assessment and management of 6 clinical cases using OSCE, based on faculty assessments; 3) Analysis of the differences in evaluation between faculty and students.

MATERIAL AND METHODS

Study Design

A practical workshop was designed within the framework of learning clinical scenarios related to abdominal pain in surgery. Six clinical scenarios pertaining to abdominal conditions were devised. Four of the scenarios were developed using a medium-fidelity simulator: obstructive ileus, paralytic ileus, liver cirrhosis, and abdominal tumor. Two additional scenarios were performed with a standardized patient: acute appendicitis and acute cholecystitis. Prior to the workshop, the students were provided with audiovisual content (a presentation and video) related to abdominal exploration.

A total of ten workshops were conducted over a 3-year period, with each workshop comprising 3 to 5 students in 6 groups (n = 670). Each student was evaluated with at least 1 checklist. Each student engages in a ten-minute OSCE station, during which they are directly observed by the faculty facilitator and the remaining students in their group. Both the students who observed the workshop and the practice teacher were responsible for the evaluation.

Measurement System

A series of specific checklists were developed for the purpose of evaluating each scenario. The items in these checklists were organized into 4 competences: communication, privacy, anamnesis, and technical skills. The number of items in each scenario's checklist varied. For instance, obstructive ileus (case 1) had 32 items, acute appendicitis (case 2) had 25 items, paralytic ileus (case 3) had 31 items, acute cholecystitis (case 4) had 24 items, liver cirrhosis (case 5) had 30 items, and abdominal tumor (case 6) had 29 items. Responses were scored as correct or incorrect for all items. The scores assigned to each clinical scenario and competency were converted into a scale of 0 to 10 for the final grade. Faculty facilitators were trained in the delivery of constructive feedback and were provided with an orientation to the new OSCE format.

Checklist Validation

Once the checklists had been drawn up, 6 experts were selected to carry out the first phase of the Delphi method, which consisted of the analysis of the different items, from which decisions were made about possible changes. Each expert provided a rating for each item on a scale of 1 to 9, with 1 representing "extremely inappropriate" and 9 corresponding to "extremely appropriate." These ratings were based on the degree to which the item was deemed appropriate for assessing the case in relation to the specific competencies of the case.

Participants

The study included 670 third-year students of the Medical Degree at the University of Murcia (UMU), Spain, as part of the subject "Medical-Surgical Skills."

Statistical Analysis

The quantitative variables were described using means and standard deviations. For group comparisons, either a t-Student test or an ANOVA test was employed, depending on the number of groups. Prior to this, the normality of the data was verified using the Shapiro-Wilk test, and the homogeneity of variances was assessed using the Levene test. In instances where a statistically significant result was obtained in the ANOVA, individual comparisons were conducted with the Bonferroni adjustment.

To assess the level of agreement among experts regarding the appropriateness of the items, the intraclass correlation coefficient (ICC) was employed. The coefficient enables the assessment of agreement in the context of a mixed-effects model comprising 2 factors. The effects of the individuals are considered random, while the effects of the measures are considered fixed. The absolute degree of agreement on the average measures is calculated. ICC values are to be interpreted as follows: Values below 0.4 indicate low reliability. Values between 0.4 and 0.75 are indicative of moderate to good reliability. Values above 0.75 are indicative of excellent reliability.

RESULTS

Validation of Checklists for Obstructive Ileus, Acute Appendicitis, Paralytic Ileus, Acute Cholecystitis, Liver Cirrhosis and Abdominal Tumor

In the first phase of the review, modifications were made by the expert panel to the following items:

-In all clinical scenarios, in the privacy competence, it was decided to separate "Identifies the patient" and "Preserves the patient's privacy" as 2 items.

-In all clinical scenarios, in the anamnesis competence, the item "Asks about characteristics of abdominal pain" was modified to "Asks about characteristics and course of abdominal pain."

-In all clinical scenarios, in the technical skills competence, the item "Describes the characteristics of the skin" was modified to "Describes the characteristics of the skin and the size of the abdomen and the size of the abdomen."

-In all clinical scenarios, in the technical skills competence, at the end, it was decided to add the item "Makes an appropriate diagnostic suspicion."

-In scenario 1 (Obstructive Ileus) and 3 (Paralytic Ileus), in the technical skills competence, the item "Describes the presence of surgical scars" was added.

-In scenario 2 (Acute Appendicitis), in the technical skills competence it was decided to separate the items "Perform a maneuver to manifest a characteristic sign of the suspected pathology" and "Perform additional maneuvers to manifest more characteristic signs of the suspected pathology" into 2 items.

-In scenarios 5 (Liver Cirrhosis) and 6 (Visible Intra-abdominal Tumor), in the technical skills competence, it was decided to keep the item "Remember the exploration of inguinal hernia points in the systematics."

In the following review, after these changes, the panel of experts did not make any further modifications to the checklists.

As we can see in [Table 1](#), we observe the high adequacy of the check-list items across all scenarios, with mean scores surpassing 7.5 points, and an ICC exceeding 0.9 in all instances.

Overall Assessment of the 6 Clinical Cases and the 4 Competences by the Teaching Staff

The students demonstrated high levels of performance across all scenarios. No significant differences were observed among the 6 scenarios. The mean scores for the scenarios were as follows: scenario 1: 8.27; scenario 2: 8.48; scenario 3: 8.13; scenario 4: 8.8; scenario 5: 7.99; and scenario 6: 8.27. Significant differences were observed in the "communication" competence, with the score for case 4 (acute cholecystitis) being significantly higher than the rest of the cases ($p = 0.005$). Similarly, significant differences were observed in the "privacy" competence between scenarios 4 (acute cholecystitis) and 6 (abdominal tumor) (10 vs 6.82; $p = 0.007$), as we can see in [Table 2](#).

Teacher-Student Discrepancies in Evaluation

In the initial 4 scenarios, no statistically significant differences were observed in the students' evaluations

TABLE 1. Validation of Checklists for Obstructive Ileus, Acute Appendicitis, Paralytic Ileus, Acute Cholecystitis, Liver Cirrhosis and Abdominal Tumor

Simulation	Experts, mean ± SD						Reliability		Strength of agreement
	1	2	3	4	5	6	ICC (CI 95%)	p-value	
Obstructive ileus	7.69 ± 1.78	7.69 ± 1.78	7.69 ± 1.78	7.54 ± 1.73	7.92 ± 1.57	7.50 ± 1.84	0.987 (0.98 – 0.99)	<0.001	Very High
Acute appendicitis	7.95 ± 0.97	7.89 ± 1.05	7.95 ± 0.97	7.84 ± 1.07	8.00 ± 0.94	7.95 ± 1.13	0.967 (0.94 – 0.99)	<0.001	Very High
Paralytic ileus	7.85 ± 0.88	7.75 ± 1.68	7.77 ± 1.55	7.64 ± 1.61	7.95 ± 1.37	7.50 ± 1.84	0.988 (0.98 – 0.99)	<0.001	Very High
Acute cholecystitis	7.92 ± 0.97	7.89 ± 1.00	7.98 ± 0.96	7.84 ± 1.07	8.02 ± 0.93	7.99 ± 1.03	0.985 (0.98 – 0.99)	<0.001	Very High
Liver cirrhosis	8.21 ± 0.83	8.21 ± 0.83	8.21 ± 0.83	8.25 ± 0.79	8.25 ± 0.94	8.25 ± 0.90	0.966 (0.94 – 0.98)	<0.001	Very High
Abdominal Tumor	8.39 ± 0.72	8.39 ± 0.72	8.39 ± 0.72	8.35 ± 0.65	8.39 ± 0.66	8.30 ± 0.88	0.950 (0.91 – 0.98)	<0.001	Very High

SD, standard deviation; ICC, intraclass correlation coefficient; CI, confidence interval.

between teachers and other students, as we can see in [Table 3](#).

In scenario 5 (liver cirrhosis), significant differences were observed between the scores given by faculty and those given by students. The overall scores were significantly higher in the student group, with a mean score of 27.97 compared to 26.5 from the faculty ($p = 0.014$). Notably, the anamnesis scores showed a significant difference, with students scoring lower at 1.87 compared to score of 2.0 given by the faculty ($p < 0.001$). The statistically significant items in this section were: "Asks about characteristics and course of abdominal pain" ($p = 0.004$) y "Asks about previous conditions that may affect the procedure" ($p = 0.002$).

Also, within scenario 5, technical skills were rated significantly higher by students, with mean scores of 21.22 versus 19.86 by the faculty ($p = 0.006$). The statistically significant items in this section were: "The examiner asks if there is any region of the abdomen that hurts more" ($p = 0.01$), "All abdominal quadrants are adequately identified" ($p = 0.001$), "Describes abnormalities in auscultation" ($p = 0.021$), "Describes dullness" ($p = 0.006$), "Palpation is initially performed superficially and then more deeply" ($p = 0.001$), "Describes hepatomegaly" ($p = 0.014$), "Describes splenomegaly" ($p = 0.042$), "The painful area is palpated last" ($p = 0.047$), y "Explores inguinal hernia points" ($p = 0.02$).

In contrast, in scenario 6 (abdominal tumor), the overall scores showed a significant difference between groups, with a mean score of 25.36 compared to 26.83 ($p = 0.026$). While communication skills did not show a significant difference (2.64 vs. 2.97, $p = 0.136$), technical skills approached significance, with scores of 18.82 versus 19.94 ($p = 0.062$). Privacy and anamnesis scores were similar between groups, with no significant differences ($p = 0.505$ and $p = 0.420$, respectively).

DISCUSSION

The evaluation of clinical cases of abdominal surgical conditions by the faculty and the students in our study revealed an important finding: a similar clinical evaluation between both groups. The students have demonstrated a high level of knowledge, not only in terms of the evaluative average by the faculty, but also by exhibiting proficiency in the evaluation of their peers. However, it is important to note that these high scores have not been compared to another measure of skill or knowledge, which limits the validation of the OSCE as the sole measure of competence. Future studies should include a comparison with other assessment methods, such as written exams, to provide a more robust evaluation.

TABLE 2. Overall Assessment of the 6 Clinical Cases And The 4 Dimensions by the Teaching Staff

	CASE						ANOVA	
	1	2	3	4	5	6	F	p-value
Total	8.27 (1)	8.48 (0.80)	8.13 (0.64)	8.80 (0.57)	7.99 (1.19)	8.27 (0.8)	1.47	0.209
Communication	8.22 (1.72)	8.22 (1.72)	8.06 (1.71)	10 (0)	7.62 (2.04)	8.79 (1.68)	3.64	0.005
Privacy	9.33 (1.76)	9.33 (1.76)	9.58 (1.44)	10 (0)	7.14 (3.78)	6.82 (4.62)	3.47	0.007
Anamnesis	9 (2.07)	10 (0)	9.58 (1.44)	9.64 (1.34)	10 (0)	10 (0)	1.58	0.177
Technical skills	8.13 (1.16)	8.32 (1.03)	7.9 (0.84)	8.49 (0.67)	7.94 (1.26)	8.18 (0.81)	0.68	0.643

The p-values in bold are statistically significant.

TABLE 3. Teacher-Student Discrepancies in Evaluation

CASE 1	Evaluator		Mean difference	t-Student	
	Faculty (n = 15)	Students (n = 224)		t	p-value
Total (0-32)	26.47 (3.2)	27.56 (3.1)	-1.09	-1.32	0.189
Communication (0-3)	2.47 (0.52)	2.67 (0.48)	-0.21	-1.62	0.108
Privacy (0-2)	1.87 (0.35)	1.84 (0.42)	0.02	0.21	0.837
Anamnesis (0-2)	1.8 (0.41)	1.89 (0.34)	-0.09	-1.02	0.311
Technical skills (0-25)	20.33 (2.9)	21.15 (2.88)	-0.81	-1.06	0.29
CASE 2					
Total (0-25)	22.53 (2.13)	23.16 (1.93)	-0.62	-1.20	0.232
Communication (0-3)	2.8 (0.41)	2.78 (0.43)	0.02	0.17	0.867
Privacy (0-2)	2 (0)	1.95 (0.21)	0.05	0.86	0.39
Anamnesis (0-2)	1.87 (0.35)	1.9 (0.31)	-0.03	-0.35	0.731
Technical skills (0-18)	15.87 (1.85)	16.53 (1.62)	-0.66	-1.52	0.131
CASE 3					
Total (0-31)	28.08 (2.94)	28.52 (2.51)	-0.44	-0.58	0.561
Communication (0-3)	2.83 (0.39)	2.9 (0.3)	-0.07	-0.75	0.451
Privacy (0-2)	2 (0)	1.96 (0.21)	0.04	0.74	0.459
Anamnesis (0-2)	1.92 (0.29)	1.93 (0.28)	-0.01	-0.11	0.91
Technical skills (0-24)	21.33 (2.5)	21.74 (2.37)	-0.41	-0.57	0.566
CASE 4					
Total (0-24)	22.5 (1.4)	22.87 (1.67)	-0.37	-0.81	0.421
Communication (0-3)	2.79 (0.43)	2.89 (0.37)	-0.11	-1.04	0.299
Privacy (0-2)	2 (0)	1.98 (0.16)	0.02	0.59	0.557
Anamnesis (0-2)	1.86 (0.36)	1.96 (0.21)	-0.10	-1.01	0.331
Technical skills (0-17)	15.86 (1.1)	16.04 (1.36)	-0.19	-0.50	0.615
CASE 5					
Total (0-30)	26.5 (2.03)	27.97 (2.15)	-1.47	-2.48	0.014
Communication (0-3)	2.64 (0.5)	2.9 (0.4)	-0.25	-1.86	0.083
Privacy (0-2)	2 (0)	1.98 (0.16)	0.03	0.59	0.554
Anamnesis (0-2)	2 (0)	1.87 (0.38)	0.13	4.85	<0.001
Technical skills (0-23)	19.86 (1.7)	21.22 (1.79)	-1.37	-2.77	0.006
CASE 6					
Total (0-27)	25.36 (2.77)	26.83 (2.04)	-1.47	-2.24	0.026
Communication (0-3)	2.64 (0.67)	2.97 (0.24)	-0.33	-1.62	0.136
Privacy (0-2)	2 (0)	1.96 (0.2)	0.04	0.67	0.505
Anamnesis (0-2)	1.91 (0.3)	1.96 (0.2)	-0.05	-0.81	0.420
Technical skills (0-22)	18.82 (1.99)	19.94 (1.91)	-1.12	-1.88	0.062

The p-values in bold are statistically significant.

Additionally, implementing an initial assessment prior to the introduction of this educational methodology would provide a clearer measure of improvement in clinical competencies.

As has been demonstrated in various studies, peer evaluation is an effective tool for fostering professionalism in medical students. Our findings align with those of other studies, which have demonstrated that peer evaluation is comparable to that conducted by educators and can be utilized as a dependable assessment tool, thus fostering active student engagement in the learning process.²⁷⁻²⁹ However, third-year medical students exhibit discrepancies in their evaluation of complex clinical cases, such as liver cirrhosis and abdominal tumor. This may be due to the greater demand for anamnesis and competencies in less common diseases. This underscores the significance of addressing specific elements, such as the identification of large masses and flat percussion in liver cirrhosis, in order to achieve an accurate diagnosis.

A detailed analysis of the results indicates that the students demonstrated a comprehensive understanding of the various competencies in all clinical scenarios. Although specific differences were observed in some dimensions, the results collectively indicate that both learning methodologies (with a simulator and a standardized patient) are effective strategies for the development of clinical skills in a patient with abdominal pain. Both methodologies are effective and complementary. Simulators are indispensable when representing certain exploratory characteristics that are not possible to represent on a standardized patient.

Previous studies have concentrated on specific outcomes related to specific conditions or have assessed student performance in a limited number of clinical scenarios, often involving standardized patients.³⁻⁵ This study, however, employs a novel approach by investigating the potential for integrating a broader range of clinical scenarios, including those that present greater complexity. In addition to evaluating technical proficiency in abdominal examination, this approach also examines the comprehensive clinical history, which is particularly beneficial in the context of an Objective Structured Clinical Evaluation (OSCE).^{10,20} Students have the opportunity to engage in constructive discourse regarding their strengths and weaknesses in the clinical development of each scenario.³⁰ This affords them the chance to observe a spectrum of physician-patient interaction styles and to hone their skills for future OSCE-type examinations.¹²⁻¹⁴

It is imperative to validate the assessment instruments to ensure the reliability and validity of the results obtained. The interobserver reliability testing addresses the issue of inter-rater variability, thereby ensuring

consistency in the interpretation and scoring of different skills.²¹ The values of the ICC of this study demonstrate a high degree of agreement among the experts on the adequacy of the items in all the scenarios, indicating the reliability and validity of the checklists used for evaluation.

It is important to acknowledge that the study did not include a baseline measure of competence prior to the implementation of the OSCE, nor did it compare the effectiveness of the OSCE with other teaching methods. This represents a limitation of the study, as we cannot definitively conclude that the observed competencies were solely due to the OSCE training. Future research should aim to include a pretest and post-test design, as well as comparisons with other educational interventions, to more accurately assess the impact of OSCE on student learning and competency development.

CONCLUSIONS

The learning method, which emphasized clinical scenarios of abdominal surgical diseases, proved to be effective in enhancing the clinical skills of third-year medical students. This approach incorporated pre-existing audiovisual materials, a hands-on workshop with medium-fidelity simulators, and standardized patients. The consistent evaluations from both students and faculty highlight the efficacy of the employed learning strategies.

However, for a more comprehensive validation, future studies should compare these results with theoretical examinations to further confirm the knowledge acquired in various competencies. Additionally, implementing an initial assessment prior to the introduction of this educational methodology would provide a clearer measure of improvement in clinical competencies. This dual approach would offer a more objective evaluation of the educational strategy's impact on the students' clinical proficiency.

REFERENCES

1. Satava RM. Surgical education and surgical simulation. *World J Surg.* 2001;25(11):1484-1489.
2. Derossis AM, DaRosa DA, Dutta S, Dunnington GL. A ten-year analysis of surgical education research. *Am J Surg.* 2000;180(1):58-61.
3. Hamm RM, Kelley DM, Medina JA, Syed NS, Harris GA, Papa FJ. Effects of using an abdominal simulator to develop palpatory competencies in 3rd year medical students. *BMC Med Educ.* 2022;22(1):1-14.

4. Juhi A, Pinjar MJ, Marndi G, Hungund BR, Mondal H. Evaluation of blended learning method versus traditional learning method of clinical examination skills in physiology among undergraduate medical students in an indian medical college. *Cureus*. 2023;15(4):e37886.
5. Gelis A, Cervello S, Rey R, Llorca G, Lambert P, Franck N, et al. Peer role-play for training communication skills in medical students: a systematic review. *Simul Healthc*. 2020;15(2):106-111.
6. Saeed S, Afzal A, Khalid F, Jehan F. Student experiences of simulation-based learning and its impact on their performance in objective structured clinical examination in pediatrics - A mixed method study. *Pak J Med Sci*. 2023;39(4):978-982.
7. Hamm RM, Kelley DM, Medina JA, Syed NS, Harris GA, Papa FJ. Effects of using an abdominal simulator to develop palpatory competencies in 3rd year medical students. *BMC Med Educ*. 2022;22(1):63.
8. Butler KL, Hirsh DA, Petrusa ER, et al. Surgery clerkship evaluations are insufficient for clinical skills appraisal: the value of a medical student surgical objective structured clinical examination. *J Surg Educ*. 2017;74(2):286-294.
9. Pugh CM, Watson A, Bell Jr RH, Brasel KJ, Jackson GP, Weber SM, et al. Surgical education in the internet era. *J Surg Res*. 2009;156(2):177-182.
10. Luo P, Shen J, Yu T, Zhang X, Zheng B, Yang J. Formative objective structured clinical examination with immediate feedback improves surgical clerks' self-confidence and clinical competence. *Med Teach*. 2023;45(2):212-218.
11. Shanks D, Wong RY, Roberts JM, Nair P, Ma IW. Use of simulator-based medical procedural curriculum: the learner's perspectives. *BMC Med Educ*. 2010;10:77.
12. Scott DJ, Cendan JC, Pugh CM, Minter RM, Dunnington GL, Kozar RA. The changing face of surgical education: simulation as the new paradigm. *J Surg Res*. 2008;147(2):189-193.
13. So HY, Chen PP, Wong GKC, Chan TTN. Simulation in medical education. *J R Coll Physicians Edinb*. 2019;49(1):52-57.
14. Lee J, Kim H, Kim KH, Jung D, Jowsey T, Webster CS. Effective virtual patient simulators for medical communication training: a systematic review. *Med Educ*. 2020;54(9):786-795.
15. Tang KS, Cheng DL, Mi E, Greenberg PB. Augmented reality in medical education: a systematic review. *Can Med Educ J*. 2020;11(1):e81.
16. Chernikova O, Heitzmann N, Stadler M, Holzberger D, Seidel T, Fischer F. Simulation-based learning in higher education: a meta-analysis. *Rev Educ Res*. 2020;90(4):499-541.
17. Takayesu JK, Farrell SE, Evans AJ, Sullivan JE, Pawlowski JB, Gordon JA. How do clinical clerkship students experience simulator-based teaching? A qualitative analysis. *Simul Healthc*. 2006;1(4):215-219.
18. Okuda Y, Bryson EO, DeMaria S Jr, Jacobson L, Quinones J, Shen B, Levine AI. The utility of simulation in medical education: what is the evidence? *Mt Sinai J Med*. 2009;76(4):330-343.
19. Nestel D, Groom J, Eikeland-Husebø S, O'Donnell JM. Simulation for learning and teaching procedural skills: the state of the science. *Simul Healthc*. 2011;6(Suppl):S10-S13.
20. Falcone JL, Watson GA. Differential diagnosis in a 3-station acute abdominal pain objective structured clinical examination (OSCE): a needs assessment in third-year medical student performance and summative evaluation in the surgical clerkship. *Mt Sinai J Med*. 2009;76(4):330-343.
21. Falcone JL, Schenarts KD, Ferson PF, Day HD. Using elements from an acute abdominal pain Objective Structured Clinical Examination (OSCE) leads to more standardized grading in the surgical clerkship for third-year medical students. *J Surg Educ*. 2011;68(5):408-413.
22. Xu X, Wang H, Luo J, Zhang C, Konge L, Tang L. Difficulties in using simulation to assess abdominal palpation skills. *BMC Med Educ*. 2023;23(1):897.
23. Omori K, Shigemoto N, Kitagawa H, Nomura T, Kaiki Y, Miyaji K, et al. Virtual reality as a learning tool for improving infection control procedures. *Am J Infect Control*. 2023;51(2):129-134.
24. O'Rourke LA, Morrison M, Grimsley A, Cotter VT. High-fidelity simulation and nurse clinical competence—an integrative review. *J Clin Nurs*. 2023;32(9-10):1549-1555.
25. Offiah G, Ekpotu LP, Murphy S, Kane D, Gordon A, O'Sullivan M, et al. Evaluation of medical student retention of clinical skills following simulation training. *BMC Med Educ*. 2019;19(1):263.

26. Jiménez López MF, Rodríguez-Conde MJ, Varela Simó G, Lozano Sánchez FS, García Criado FJ, Novoa Valentín NM, et al. Validación de los sistemas de evaluación de un programa de enseñanza práctica de las habilidades clínicas y quirúrgicas en las asignaturas de Fundamentos de Cirugía de los grados de Medicina y Odontología. 2012.
27. Spandorfer J, Puklus T, Rose V, Vahedi M, Collins L, Giordano C, et al. Peer assessment among first year medical students in anatomy. *Anat Sci Educ*. 2014;7(2):144-152.
28. DuCoin C, Zuercher H, McChesney SL, Korndorffer JR Jr.. Peer assessment in medical student education: a study of feasibility, benefit, and worth. *Am Surg*. 2022;88(9):2361-2367.
29. Yu JH, Lee MJ, Kim SS, Yang MJ, Cho HJ, Noh CK, et al. Assessment of medical students' clinical performance using high-fidelity simulation: comparison of peer and instructor assessment. *BMC Med Educ*. 2021;21(1):506.
30. Febrero B, Almela-Baeza J, Ros I, et al. The impact of information and communications technology and broadcasting on YouTube for improving attitude toward organ donation in secondary education with the creation of short films. *Patient Educ Couns*. 2021;104(9):2317-2326.