



ORIGINAL ARTICLE

Competency assessment of residents of Intensive Care Medicine through a simulation-based objective structured clinical evaluation (OSCE). A multicenter observational study



Á. Castellanos-Ortega^a, M.J. Broch^a, D. Palacios-Castañeda^b, V. Gómez-Tello^c, M. Valdivia^b, C. Vicent^a, I. Madrid^a, N. Martínez^b, M.J. Párraga^d, E. Sancho^a, M.d.C. Fuentes-Dura^e, M.D. Sancerni-Beitia^e, R. García-Ros^{f,*}

^a Intensive Care Department, Hospital Universitario y Politécnico La Fe, Valencia, Spain

^b Intensive Care Department, Hospital Puerta de Hierro-Majadahonda, Spain

^c Intensive Care Department, University Hospital Moncloa, Madrid, Spain

^d Intensive Care Department, Hospital Universitario Morales Meseguer, Murcia, Spain

^e Department of Methodology of the Behavioral Sciences, University of Valencia, Spain

^f Department of Developmental and Educational Psychology, University of Valencia, Spain

Received 7 December 2021; accepted 22 January 2022

KEYWORDS

Training;
Assessment;
Simulation;
Competency-based
education;
Intensive care

Abstract

Objectives: The current official model of training in Intensive Care Medicine (ICM) in Spain is based on exposure to experiences through clinical rotations. The main objective was to determine the level of competency (I novice to V independent practitioner) achieved by the residents at the end of the 3rd year of training (R3) in ICM through a simulation-based OSCE. Secondary objectives were: (1) To identify gaps in performance, and (2) To investigate the reliability and feasibility of conducting simulation-based assessment at multiple sites.

Design: Observational multicenter study.

Setting: Thirteen Spanish ICU Departments.

Participants: Thirty six R3.

Intervention: The participants performed on five, 15-min, high-fidelity crisis scenarios in four simulation centers. The performances were video recorded for later scoring by trained raters.

Main variables of interest: Via a Delphi technique, an independent panel of expert intensivists identified critical essential performance elements (CEPE) for each scenario to define the levels of competency.

* Corresponding author.

E-mail address: rafael.garcia@uv.es (R. García-Ros).

Results: A total of 176 performances were analyzed. The internal consistency of the check-lists were adequate (KR-20 range 0.64–0.79). Inter-rater reliability was strong [median Intraclass Correlation Coefficient across scenarios: 0.89 (0.65–0.97)]. Competency levels achieved by R3 were: Level I (18.8%), II (35.2%), III (42.6%), IV/V (3.4%). Overall, a great heterogeneity in performance was observed.

Conclusion: The expected level of competency after one year in the ICU was achieved only in half of the performances. A more evidence-based educational approach is needed. Multiple center simulation-based assessment showed feasibility and reliability as an evaluation method of competency.

Trial registration: COBALIDATION. NCT04278976. (<https://register.clinicaltrials.gov>).

© 2022 The Author(s). Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

PALABRAS CLAVE

Formación;
Evaluación;
Simulación;
Formación basada en competencias;
Medicina intensiva

Evaluación del nivel de competencia de residentes de Medicina Intensiva mediante una evaluación clínica, objetiva y estructurada basada en simulación (ECOE). Estudio multicéntrico observacional

Resumen

Objetivos: El modelo de formación en medicina intensiva (MI) en España se basa en la experiencia adquirida durante una serie de rotaciones programadas por diferentes áreas clínicas. El objetivo principal del estudio fue determinar el nivel de competencia (I principiante – V autónomo) de los residentes de MI al finalizar el tercer año de residencia (R3) mediante una ECOE basada en simulación. Objetivos secundarios: 1) identificar brechas en el desempeño; 2) investigar la fiabilidad y validez de una ECOE simulada multicéntrica como método de evaluación.

Diseño: Estudio multicéntrico observacional.

Ámbito: Trece servicios españoles de Medicina Intensiva.

Participantes: Treinta y seis R3.

Intervención: Los 36 R3 participaron en cinco escenarios clínicos simulados de 15 minutos de duración en cuatro centros de simulación. Las actuaciones se grabaron en video y posteriormente se calificaron por pares de expertos.

Variables de interés principales: Un panel de intensivistas expertos seleccionó mediante el método Delphi los elementos críticos esenciales de cada escenario para definir los niveles de competencia.

Resultados: La consistencia interna de los listados de verificación fue adecuada (KR-20:0,64–0,79). La fiabilidad interjueces fue elevada (coeficiente de correlación intraclass [mediana]: 0,89 [0,65–0,97]). Los niveles de competencia conseguidos fueron: nivel I (18,8%), II (35,2%), III (42,6%), IV/V (3,4%). Globalmente, se observó una gran heterogeneidad en el desempeño.

Conclusión: El nivel de competencia esperado se logró únicamente en la mitad de las actuaciones. Se necesita un modelo de formación más basado en objetivos y evidencias. La evaluación mediante escenarios simulados en múltiples centros demostró ser factible y fiable.

© 2022 El Autor(s). Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

The current postgraduate medical training model in most European countries is the so called “time-based training”.^{1–3} This paradigm assumes that mere exposure to clinical experiences based on temporary rotations through different Departments suffices to acquire the necessary professional competencies. Opportunistic learning and volume of practice rather than learning guided by objectives defines competency. Certification depends on a logbook of cases, and a generic, subjective report about knowledge, technical and non-technical skills acquired by the resident, which is

performed after every rotation and yearly by the tutors in charge. A knowledge-based examination is also included in some countries. At present, some prestigious national educational institutions are transitioning from a time-based to a competency-based medical education (CBME) system which is a learner-centered approach that emphasizes achieving specific outcomes.^{4–8} The CBME model for ICM in Europe is called CoBaTrICE (Competency Based Training in Intensive Care Medicine in Europe).^{9–14} The implementation of CBME is challenging because requires organizational changes, resources, particularly more dedication of teaching time, as well as the training of tutors and staff members

in formative assessment and feedback techniques.^{15,16} Research in this field is still limited because the model has been applied in a partial and scantily structured manner. In order to evaluate whether the implementation of CoBaTrICE provides higher levels of competency in comparison with the current official time-based program in ICM in Spain, a multicentric cluster-based randomized trial is currently ongoing.¹⁷ Before starting the implementation of CoBaTrICE we performed a baseline observational study with the primary objective of determining the level of competency achieved by the residents at the end of the 3rd year of training in ICM. We chose a simulation-based objective structured clinical evaluations (OSCE) to assess the ability to integrate knowledge, judgment, communication, and teamwork into the simulated practice setting.^{18–23} Secondary objectives were: (1) to identify gaps in performance that could be addressed in future educational interventions, and (2) to investigate the reliability and feasibility of conducting simulation-based assessment at multiple sites.

Methods

Design and setting

ICM in Spain is a five-year primary specialty divided in two stages: Stage 1 consists of an initial two-year block (R1–R2) of training that is spent in anesthesia and medicine; Stage 2 consists of a three-year block (R3–R5) that covers general and specific ICM training in a variety of “special” areas including coronary care, polytrauma, pediatric, neurosurgical, post-transplant and cardiothoracic ICM.²⁴

We conducted an observational multicenter study to determine the performance of trainees at the end of R3, actually at the end of the first full year working fulltime at the ICU, through a simulation-based OSCE.

Participants

There were 36 R3 consent participants belonged to 13 ICU Department’s from 13 academic referral hospitals located in Spain. The participating ICUs are general medical and surgical ICUs accredited to train 2–3 new residents in ICM per year.

Several socio-educational variables of the participants were recorded: age, gender, grade point average (GPA) at medical school, MIR entrance exam position, previous simulation-based training experience, and the size of the hospital where they were performing the residency.

The study was approved by the ethics committee of the Instituto de Investigación Sanitaria La Fe and registered to Clinical Trials (ClinicalTrials.gov NCT04689477). After obtaining informed consent, participants who volunteered for the study were allocated to the simulation scenarios.

Intervention

The OSCE was performed in April and May 2019 at four simulation centers geographically close to participant hospitals: Hospital la Fe, Valencia; Hospital Clinic, Barcelona; IAVANTE,

Granada; University Francisco de Vitoria, Madrid. Each participant performed in five 15-min, standardized patient or high-fidelity simulated clinical crisis scenarios.

Designing five standardized scenarios and rating instruments

Via a Delphi technique, an independent panel of 10 intensivists subject matter experts [simulation instructors and/or European Diploma Intensive Care (EDIC) examiners] performed the following tasks: (1) To select the CoBaTrICE competences to be assessed according with the level of training of the participants. (2) To design the scenarios; the five scenarios that were approved for use by consensus were: (1) management of septic shock, ARDS, and endotracheal intubation; (2) neurocritical care and intra-hospital transport; (3) acute coronary syndrome management and cardiopulmonary resuscitation; (4) postoperative management, hemorrhagic shock; (5) Initial assessment and management of the multiple-trauma patient. (3) To define the items of the checklist for each scenario; each checklist included 20–25 items that were classified as follows: (a) the critical essential performance elements (CEPE), and (b) the critical non-essential performance elements (CNEPE) that must be observed and scored in a yes/no format. CEPEs are defined as essential steps or actions in the management of the patient which if missed could have an immediate significant impact on morbidity and mortality. CNEPEs are also important for the adequate management of the patient but they don’t have an immediate influence on the outcome. There were 7–12 CEPEs and 13–15 CNEPEs in each scenario (see the check list and scoring system of scenario number 1 in [additional file 1](#)).

The performances were video recorded.²⁵ The videos were randomly assigned and then rated by two blinded raters, members of the experts panel using the specific checklists with a detailed description of the competencies technical (diagnosis and treatment) and non-technical (communication, team leadership, resource management) associated with each item. After each video-assessment, the performance of the resident was classified in a level of competency on a descriptive scale of I–V ([Table 1](#)).

Measures included: (1) the percentage of CEPEs observed; (2) the percentage of CNEPEs observed; (3) the competency level achieved in each scenario; (4) the total scoring achieved in each scenario which was calculated as follows:

Total score (range 0–100) = (number of CEPE completed $\times$ 2 points + number of CNEPE completed $\times$ 1 point) $\times$ 100/potential maximum score achievable.

Standardization of scenario delivery^{20,26}

The design of the scenarios involved the use of “high fidelity mannekins” (Meti HPS® and iStan®), and also “standardized patient actors”. In order to facilitate reproducible scenario delivery rules, detailed scripts and a guidebook for each scenario were created. Participants were briefed on relevant mannekin characteristics, clinical equipment, and other resources. Confederates played the role of nurse, senior intensivist, surgeon, anesthesiologist, radiologist, relatives, etc. Each participant performed in the five different simulation scenarios as primary intensivist.

Table 1 Levels of competence considered in the study.

Level	CEPES/CNEPES performed appropriately	Level of autonomy/support needs
I	Less than 60% of the CEPES	The participant needs guidance and direct supervision in all cases.
II	Equal or more than 60% but less than 80% of the CEPES	The participant needs guidance and supervision in most of situations.
III	Equal or more than 80% but less than 100% of CEPES	The participant needs some guidance and supervision in complex situations.
IV	All CEPES (100%) but less than 60% of CNEPES	The participant can perform the activity under indirect supervision.
V	All CEPES (100%) and $\geq 80\%$ of the CNEPES.	The participant is independent to perform the activity.

CEPE: critical essential performance elements; CNEP: critical non-essential performance elements.

In order to check the time, difficulty, feasibility and reliability of the scenarios as well as the video rating process, a pilot OSCE with three R3 non-participant in the study was carried out at the simulation Center in Hospital La Fe, Valencia, Spain.

Following the simulation performance, the participants were asked to complete a questionnaire regarding their perceived fidelity of their scenarios to 'real-life' and their opinion of the utility and suitability of their scenarios for competency examinations.

Main outcomes

Primary outcome

The level of competency achieved by the R3 at the end of the first year of specific training in ICM defined as the percentage of scenarios assessed through the simulation-based OSCE in which level III or higher was achieved.

Secondary outcomes

1. Total scoring.
2. Percentage of CEPES completed.
3. Percentage of CNEPES completed.

Statistical analysis

This study represents baseline OSCE findings of the ongoing Cobalidation trial. A statistical power analysis was performed to determine the minimum sample size required to conduct the study with a statistical power of .95, setting the Type I error at the standard cut-off value ($\alpha = .05$) and a medium-large effect size ($f = .31$), considering two experimental conditions (competency-based training program vs. traditional training program). Results indicated that a minimum sample size of 72 observations (36 residents) was required.

Results are shown as median, Interquartil Range (IQR) and maximum–minimum range.

To compare continuous variables, Kruskal–Wallis test were used as appropriate. To compare categorical variables, the Chi-Square test was used. All tests were two-tailed, and $p < .05$ was predetermined to define statistical significance. Internal consistency reliability of the check-lists was

calculated by the Kuder–Richardson coefficient 20 (KR-20). Inter-rater reliability for the OSCE scenarios was estimated with Fleiss kappa and the Intraclass Correlation Coefficient (ICC).

All Analysis were performed using SPSS statistical package version 23.0 (SPSS Inc., Chicago, IL).

Results

A total of 36 R3 from 13 ICU Departments of 13 Spanish teaching hospitals performed in 176 scenarios. One hospital with two residents eventually declined to participate (Table 2).

Six video-records of scenarios one (1), scenario two (2) and scenario three (3) were lost. The distribution of the participants in the OSCE in each of the four simulation centers was as follows: Barcelona, 9; Madrid, 13; Granada, 6; and Valencia, 8. There were 12 male and 24 female, mean age 29.8 ± 4.4 (range 27–49 years). Half of them had obtained an average grade in their undergraduate studies equal to or higher than B, and half of them had obtained a MIR position lower than number 3000 (range 110–4534). Regarding their previous experience in simulation-based training, 79% had taken a course in advanced cardiopulmonary resuscitation, 25% a course of the management of polytrauma patients, 14% a course of acute crisis resource management, and 72% other simulation-based courses. Fourteen of them were doing their residency in hospitals with more than 1000 beds (five hospitals), fifteen in hospitals with 600–1000 beds (five hospitals), and seven of them in hospitals with less than 600 beds (three hospitals).

Check list internal consistency, inter-rater and intra-rater reliability

The internal consistency of the check-lists created ad hoc to assess residents' performance in the OSCE scenarios is shown in Table 3. The KR-20 coefficients ranged between 0.642 (Acute coronary syndrome) and 0.791 (septic shock, ARDS and endotracheal intubation). Regarding interrater reliability, Fleiss' Kappa across scenarios ranged between 0.570 (septic shock, ARDS and endotracheal intubation) and 0.871 (Post-operative management). ICC ranged from 0.560 (Acute coronary syndrome) to 0.871 (Multiple-trauma patient).

Table 2 Hospitals participants (ICU Departments), number of residents/year accredited, number of residents participants in the study and number of beds of the hospitals.

Hospitals	Residents/year N	Residents participants N	Number of beds
Consorti Corporació Sanitària Parc Taulí, Sabadell, Barcelona.	3	3	583
Hospital Clínico Universitario de Valencia, Valencia.	2	2	582
Hospital Clínico San Carlos, Madrid.	3	3	861
Hospital Clínico Universitario Virgen de la Arrixaca, Murcia.	3	3	920
Hospital General Universitario de Alicante, Alicante.	3	3	825
Hospital Universitario Doctor Peset, Valencia.	2	2	539
Hospital Universitario 12 de Octubre, Madrid.	3	3	1256
Hospital Universitario Germans Trias i Pujol, Badalona, Barcelona.	3	3	516
Hospital Universitario de Gran Canaria Doctor Negrín. Las Palmas de Gran Canaria.	3	3	621
Hospital Universitario La Paz, Madrid.	3	2	1308
^a Hospital Universitario Virgen de la Macarena, Sevilla.	2	0	866
Hospital Universitario Virgen de la Nieves, Granada.	3	3	918
Hospital Universitario Virgen del Rocío, Sevilla.	3	3	1251
Hospital Universitario Vall d'Hebron, Barcelona.	3	3	1146

^a Finally declined to participate.

In the post-assessment survey the participants “strongly agreed” the scenarios were realistic, the duration was appropriate, and the competences assessed were relevant to their clinical practice (Table 4).

Residents’ performance

Thirty performances (17%) needed a third rater due to a discrepancy in two or more CEPEs completed between the two raters. The results of the residents’ performance are shown in Table 5. The median score of the participants across the 176 performances in the five crisis management scenarios was 71 points (IQR 63–78). The median of CEPEs completed was 77.5% (IQR 64–86). In 54% of the performances, at least three CEPEs were missed. The highest percentage of CEPEs completed was observed in the scenario “acute coronary syndrome management and cardiopulmonary resuscitation”: 85% (IQR 70–81). The levels of competence achieved in the 176 performances were: Level I: 33 (18.8%); level II: 62 (35.2%); level III: 75 (42.6%); levels IV and V were exceptionally achieved, 5 (2.8%) and 1 (0.6%) respectively. The level III was more often achieved in the scenario “acute coronary syndrome management and cardiopulmonary resuscitation” (82%), followed by “postoperative management” (53%), “Multiple-trauma patient” (44%), “neurocritical care” (38%), and “septic shock, ARDS and endotracheal intubation” (17%). Overall, a great heterogeneity was observed regarding the scores obtained in the various scenarios by the residents (Table 6) and hospitals (additional file 2, Tables 6S and 6Sa–6Se).

Differences in performance among simulation centers

There were no statistically significant differences among the residents’ level of competency achieved in the various simulation centers (Chi-Square test, $p = .25$).

Residents’ socio-demographic and educational variables

Except in scenario 3 where female performed better than male, there were not significant differences between the level of competence achieved by gender, GPA obtained in the medical degree, MIR enter examination position, previous experience in simulation-based training, and the hospital size (number of beds). Results are shown in additional file 3 (Tables 7Sa–7Se).

Discussion

Since traditional examinations fail to capture the uncertainty that will be encountered in some clinical scenarios and the assessment of advance skills is difficult in real practice, we created a simulated-based OSCE which was reliable and reproducible across four advanced simulation centers. The total median score achieved by the 36 R3 participants in a total of 176 performances was 71 out of 100 points. The percentage of CEPEs accomplished was 75%. However, only half of the participants achieved the expected level of competency. The experts panel considered that level III (the participant needs supervision to perform the activity in complex situations) should be the appropriate level of competency for that specific stage of training. However, level III was successfully achieved only in the scenario “acute coronary syndrome”. The worst results were obtained in the scenario “management of the septic shock, ARDS and endotracheal intubation”, here, most of the participants applied correctly the 1-hour sepsis bundle and protective mechanical ventilation strategy, however, the approach to the patient was disorganized, and they failed to apply a complete protocol of endotracheal intubation (Additional file 1, table* 1S) in a high-risk patient with severe hypoxemia and shock.²⁷ There was also a room for improvement in the so called non-technical skills such as leadership, setting

Table 3 Internal consistency. Reliability of the check-lists determined by Kuder-Richardson coefficient 20, and Inter-rater reliability determined by Fleiss KAPPA coefficient for the dichotomic items of the check lists and Intraclass Correlation Coefficient for the scores given by the raters.

Scenario	Internal consistency KR-20	Inter-rater reliability	
		Kappa (95 CI)	ICC (95 CI)
1. Management of septic shock, ARDS and endotracheal intubation.	0.791	0.570 (0.500–0.640)	0.680 (0.309–0.852)
2. Neurocritical care and intra-hospital transport.	0.660	0.632 (0.571–0.693)	0.832 (0.640–0.921)
3. Acute coronary syndrome management and cardiopulmonary resuscitation.	0.642	0.585 (0.510–0.660)	0.560 (0.200–0.803)
4. Postoperative management, hemorrhagic shock.	0.681	0.653 (0.585–0.721)	0.836 (0.652–0.923)
5. Initial assessment and management of the multiple-trauma patient.	0.696	0.643 (0.582–0.704)	0.871 (0.726–0.940)

KR-20: Kuder-Richardson coefficient 20; ICC: Intraclass Correlation Coefficient; CI: confidence interval.

Table 4 Residents' feedback on the simulation based OSCE reported as scoring of the level of agreement to the OSCE from a five-point Likert scale (1 = strongly disagree; 5 = strongly agree).

Statement	Mean	SD	Min	Max
1. The scenarios included many relevant competences I need to practice intensive care medicine	4.62	0.49	4.00	5.00
2. The number of scenarios is enough to assess the most important competencies	4.24	0.74	2.00	5.00
3. The design of the scenarios was adequate	4.59	0.50	4.00	5.00
4. The organization of the OSCE was adequate	4.71	0.52	3.00	5.00
5. Simulation should be used as one of several assessment modalities during my residency	4.65	0.54	3.00	5.00
6. Simulation is an appropriate tool to assess management of crisis in intensive care medicine	4.62	0.65	3.00	5.00
7. The scenario was realistic	4.53	0.61	3.00	5.00
8. I would recommend this experience to others	4.68	0.59	3.00	5.00

priorities, and communication with patients and relatives. Although there are not similar studies performed in ICM, our results are not very different from those obtained in studies conducted to assess performance of anesthesia professionals using simulation.^{19,28} Weinger et al.²² in their study of 268 board-certified anesthesiologists found that CEPEs were commonly omitted. Approximately 30% of encounters were rated as "poor" or as "unsatisfactory" for overall individual technical or behavioral performance. They documented omissions, errors, or delays in actions considered by clinical experts to be critical to successful patient care. As in other studies,^{19,22,28,29} a wide variability in performance was found. In general, the same resident performed significantly different in some scenarios in comparison to others, there were also differences in performance among the residents of the same hospital. This observed heterogeneity might be due to the characteristics of our traditional experience-based training model where assessments of residents are scarce, indirect and subjective. In this context, it is hard to ensure that residents have the expertise they need to perform entrusted tasks. The CBME model proposes more solid principles such as defining the specific learning outcomes and focusing attention upon the development and

demonstration of skills, attitudes and knowledge acquired by residents during the training process. Frequent formative work-based assessments as well as the record of the learning experiences in a portfolio are essential elements to promote learning, self-reflection, progression, and ultimately to guarantee that the predefined competences and skills are effectively acquired.^{5,30,31}

Study limitations

Although the OSCE does provide a standardized and relatively objective method of evaluating a set of clinical skills in medicine, its use does not guarantee accurate decisions about examinees, especially, when referring to the non-technical skills,³² multiple relevant factors such as the number of scenarios, items, scoring, examiners, etc., can influence the results. In addition, although simulation is being widely used to measure technical and non-technical skills^{21,33–37} there is little documentation of a relationship between simulation performance and performance in the clinical setting.³⁸ It has just been shown that those with more training and experience perform better in the

Table 5 Total scoring, percentage of CEPE and CNEPE completed, and competency level frequency achieved in the 176 performances in the five OSCE scenarios. Results are shown as median, Interquartil Range (IQR) and maximum-minimum range.

Scenario	Performances N	Total scoring			CEPE completed (%)			CNEPE completed (%)			Competency level N performances				
		Median	IQR	Range	Median	IQR	Range	Median	IQR	Range	I	II	III	IV	V
1. Management of septic shock, ARDS and endotracheal intubation.	35	70	62–77	43–89	64	57–71	36–86	76	63–82	49–96	14	15	6	0	0
2. Neurocritical care and intra-hospital transport.	36	67	52–71	32–84	72	59–81	31–100	57.5	46–64	16–85	11	12	12	1	0
3. Acute coronary syndrome management and cardiopulmonary resuscitation.	33	77	70–81	47–89	85	80–90	55–100	62	55–72	21–87	1	5	25	2	0
4. Postoperative management, hemorrhagic shock.	36	72.5	65–84	46–97	80	70–90	42–100	61	53–69	27–91	2	15	16	2	1
5. Initial assessment and management of the multiple-trauma patient.	36	67.5	58–78	41–91	78	68–87	41–97	59	47–64	19–83	5	15	16	0	0
Overall ratings	176	71	63–78	32–97	77.5	64–86	31–100	62	52–72	16–96	33	62	75	5	1

CEPE: critical essential performance elements; CNEP: critical non-essential performance elements.

Table 6 Correlations between Total Score resident' scores in the various OSCE scenarios.

Scenarios	1	2	3	4
1. Management of septic shock, ARDS and endotracheal intubation.				
2. Neurocritical care and intra-hospital transport.	.27			
3. Acute coronary syndrome management and cardiopulmonary resuscitation.	.42*	.07		
4. Postoperative management, hemorrhagic shock.	-.00	.22	.08	
5. Initial assessment and management of the multiple-trauma patient.	.33	.35*	.05	.64***
Pearson correlation coefficient (* $p < .05$, ** $p < .01$; *** $p < .001$).				

scenarios, suggesting that simulation-based assessments may ultimately prove useful as an indicator that they are ready for unsupervised practice in the real world.^{19,23,29,39}

We designed five 15 min clinical scenarios based on the model "crises resource management",^{18,40} where rapid decision-making based on incomplete data is needed. Each participant performed as primary intensivist and worked in a team with trained confederate clinicians and nurses to provide an environment as much realistic as possible. Although the simulated clinical environment was not identical to the participants usual one, where they would probably perform better, they "strongly agreed" in the post-assessment survey the scenarios were realistic, the duration was appropriate, and the competences assessed were relevant to their clinical practice, what further supports construct-related validity. Like in other studies, previous experience in simulation did not influence the level of competency achieved for the residents.^{19,22} Finally, the definitions of the levels of performance were established by a consensus of a panel of experts, thus, they are debatable.

Despite these limitations, we still think this study provides a pathway to identify gaps in performance in common problem areas.

Conclusions

The expected level of competency after a year in the ICU was achieved only in half of the performances. Gaps were observed in compliance with evidence-based protocols and also in non-technical skills. Reliance on the traditional experience-based training model alone might be insufficient for ensuring quality and safety in patient care. A more evidence-based educational approach is needed.

Multiple center simulation-based assessment showed feasibility, validity and reliability as an evaluation method of competency of the residents in the scenarios. It could be replicated for formative assessments, and even for nationwide comparisons and performance benchmarking. Additional research is needed to determine how simulation-based assessments predict performance in clinical settings.

Authors' contributions

ACO and RGR had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. ACO and RGR were responsible for the study design, data analysis, and interpretation, and the writing of the manuscript. MCFD and MDSB were responsible for the statistical analysis. ACO, MJB, DPC, VGT, MV, CV,

IM, NM, ES, MJP were responsible for selection the CoBa-TrICE competences to be assessed, design the scenarios, define the items of the checklist for each scenario, organization and implementation of the OSCE in the four simulation centers, assessment and rating the performances. ACO was responsible for participants recruitment. All authors read and approved the final manuscript.

Funding

The study has received funding from:

1. ESICM Trials Group Award 2018. TRIALS GROUP AWARDS. European Society of Intensive Care Medicine.
2. Conselleria de Educació, Investigació, Cultura y Deporte. Generalitat Valenciana. Código proyecto AICO/2018/126.

Conflict of interests

The authors declare that they have no competing interests.

Acknowledgements

The tutors and heads of the ICU Departments from: Hospital Universitario la Paz, Madrid. Abelardo García de Lorenzo, Maria José Asensio. Hospital Clínico San Carlos, Madrid: Miguel Sánchez, Manuel Álvarez. Hospital Universitario 12 de octubre, Madrid: Juan Carlos Montejo, José Luis Pérez Vela. Hospital Clínico Universitario Virgen de la Arrixaca, Murcia: Rubén Jara, Carlos Luis Albacete. Hospital Universitario Virgen de las Nieves, Granada: José Miguel Pérez Villares, Maria Redondo. Hospital Universitario Virgen del Rocío, Sevilla: Rosario Amaya, Yael Corcia. Hospital Universitario de Gran Canaria doctor Negrin, Las Palmas De Gran Canaria: Sergio Ruiz, Catalina Sanchez. Hospital Clínico Universitario, Valencia: Marisa Blasco, Angela Jorda. Hospital Universitario Doctor Peset, Valencia: Rafael Zaragoza, Santiago Borrás. Hospital General Universitario, Alicante: Francisco Jaime, José Luis Anton. Consorci Corporació Sanitària Parc Taulí, Sabadell: Ana Ochagavía, Ana Navas. Hospital Vall D'hebron, Barcelona: Ricard Ferrer, Marcos Perez. Hospital Universitari Germans Trias i Pujol Badalona: Pilar Ricart, Fernando Armestar.

Simulation centers: Hospital la Fe, Valencia; Hospital Clinic, Barcelona; IAVANTE, Granada; University Francisco de Vitoria, Madrid.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.medint.2022.01.011](https://doi.org/10.1016/j.medint.2022.01.011).

References

- Bion J, Rothen HU. Models for intensive care training. A European perspective. *Am J Respir Crit Care Med*. 2014;189:256–62, <http://dx.doi.org/10.1164/rccm.201311-2058CP>.
- Jonker G, Manders LA, Marty AP, Kalkman CJ, Ten Cate TJ, van Gessel EF, Hoff RG. Variations in assessment and certification in postgraduate anaesthesia training: a European survey. *Br J Anaesth*. 2017;119:10, <http://dx.doi.org/10.1093/bja/aex196>.
- Kjaer NK, Kodal T, Shaughnessy AF, Qvesel D. Introducing competency-based postgraduate medical training: gains and losses. *Int J Med Educ*. 2011;2:110–5, <http://dx.doi.org/10.5116/ijme.4e78.427f>.
- Carraccio C, Englander R, Van Melle E, Ten Cate O, Lockyer J, Chan MK, et al. Advancing competency-based medical education: a charter for clinician-educators. *Acad Med*. 2016;91:645–9, <http://dx.doi.org/10.1097/ACM.0000000000001048>.
- Frank JR, Snell L, Englander R, Holmboe ES, Collaborators I. Implementing competency-based medical education: moving forward. *Med Teach*. 2017;39:568–73, <http://dx.doi.org/10.1080/0142159X.2017.1315069>.
- Competency by design. Royal College of Physicians and Surgeons of Canada; 2018. <https://www.royalcollege.ca/rcsite/cbd/competence-by-design-cbd-e> [accessed 10.11.21].
- The Faculty of Intensive Care Medicine. <https://www.ficm.ac.uk/trainingcurricula-and-assessment/core-and-specialist-training> [accessed 10.11.21].
- Accreditation Council for Graduate Medical Education (ACGME). <https://www.acgme.org/What-We-Do/Accreditation/Milestones/Resources> [accessed 10.11.21].
- CoBaTrice: <https://www.esicm.org/education/cobatrice/> [accessed 01.12.21].
- Bion JF, Barrett H. Development of core competencies for an international training programme in intensive care medicine. *Intensive Care Med*. 2006;32:1371–83, <http://dx.doi.org/10.1007/s00134-006-0215-5>. PMID: 16841214.
- CoBaTrICE Collaboration. The views of patients and relatives of what makes a good intensivist: a European survey. *Intensive Care Med*. 2007;33:1913–20, <http://dx.doi.org/10.1007/s00134-007-0799-4>.
- CoBaTrICE Collaboration. The educational environment for training in intensive care medicine: structures, processes, outcomes and challenges in the European region. *Intensive Care Med*. 2009;35:1575–83, <http://dx.doi.org/10.1007/s00134-009-1514-4>.
- CoBaTrICE Collaboration. International standards for programmes of training in intensive care medicine in Europe. *Intensive Care Med*. 2011;37:385–93, <http://dx.doi.org/10.1007/s00134-010-2096-x>.
- Roca J, Pérez JM, Colmenero M, Muñoz H, Alarcón L, Vázquez G. Professional competence for the care of critical patients: beyond specialties. *Med Intensiva*. 2007;31:473–84, [http://dx.doi.org/10.1016/S0210-5691\(07\)74854-4](http://dx.doi.org/10.1016/S0210-5691(07)74854-4).
- Castellanos-Ortega A, Rothen HU, Franco N, Rayo LA, Martín-Loeches I, Ramírez P, et al. Training in Intensive Care Medicine. A challenge within reach. *Med Intensiva*. 2014;38:305–10, <http://dx.doi.org/10.1016/j.medint.2013.12.011>.
- Caverzagie KJ, Nousiainen MT, Ferguson PC, Ten Cate O, Ross S, Harris KA, et al. Overarching challenges to the implementation of competency-based medical education. *Med Teach*. 2017;39:588–93, <http://dx.doi.org/10.1080/0142159X.2017.1315075>.
- Effect of a Competence Based Medical Education Programme on Training Quality in Intensive Care Medicine. (COBALIDATION). NCT04278976. (<https://register.clinicaltrials.gov>).
- Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology*. 1998;89:8–18, <http://dx.doi.org/10.1097/0000542-199807000-00005>.
- Murray DJ, Boulet JR, Avidan M, Kras JF, Henrichs B, Woodhouse J, Evers AS. Performance of residents and anesthesiologists in a simulation-based skill assessment. *Anesthesiology*. 2007;107:705–13, <http://dx.doi.org/10.1097/01.anes.0000286926.01083.9d>.
- Hastie MJ, Spellman JL, Pagano PP, Hastie J, Egan BJ. Designing and implementing the objective structured clinical examination in anesthesiology. *Anesthesiology*. 2014;120:196–203, <http://dx.doi.org/10.1097/ALN.0000000000000068>.
- Chiu M, Tarshis J, Antoniou A, Bosma TL, Burjorjee JE, Cowie N, et al. Simulation-based assessment of anesthesiology residents' competence: development and implementation of the Canadian National Anesthesiology Simulation Curriculum (CanNASC). *Can J Anaesth*. 2016;63:1357–63, <http://dx.doi.org/10.1007/s12630-016-0733-8>.
- Weinger MB, Banerjee A, Burden AR, McIvor WR, Boulet J, Cooper JB, et al. Simulation-based assessment of the management of critical events by board-certified. *Anesthesiol*. 2017;1:475–89, <http://dx.doi.org/10.1097/ALN.0000000000001739>.
- Boyle WA, Murray DJ, Beyatte MB, Knittel JG, Kerby PW, Woodhouse J, Boulet JR. Simulation-based assessment of critical care "front-line" providers. *Crit Care Med*. 2018;46:e516–22, <http://dx.doi.org/10.1097/CCM.0000000000003073>.
- Junta Directiva de la Sociedad Española de Medicina Intensiva Crítica y Unidades Coronarias (SEMICYUC). Medicina intensiva en España [Intensive medicine in Spain]. *Med Intensiva*. 2011;35:92–101, <http://dx.doi.org/10.1016/j.medint.2010.12.008>.
- Easdown LJ, Wakefield ML, Shotwell MS, Sandison MR. A Checklist to help faculty assess ACGME milestones in a video-recorded OSCE. *J Grad Med Educ*. 2017;9:605–10, <http://dx.doi.org/10.4300/JGME-D-17-00112.1>.
- McIvor WR, Banerjee A, Boulet JR, Bekhuis T, Tseytlin E, Torsher L, et al. A taxonomy of delivery and documentation deviations during delivery of high-fidelity simulations. *Simul Healthc*. 2017;12:1–8, <http://dx.doi.org/10.1097/SIH.0000000000000184>.
- Russotto V, Myatra SN, Laffey JG, Tassistro E, Antolini L, Bauer P, et al. Intubation practices and adverse peri-intubation events in critically ill patients from 29 countries. *JAMA*. 2021;325:1164–72, <http://dx.doi.org/10.1001/jama.2021.1727>.
- Blum RH, Boulet JR, Cooper JB, Muret-Wagstaff SL, Harvard Assessment of Anesthesia Resident Performance Research Group. Simulation-based assessment to identify critical gaps in safe anesthesia resident performance. *Anesthesiology*. 2014;120:129–41, <http://dx.doi.org/10.1097/ALN.0000000000000055>.
- Blum RH, Muret-Wagstaff SL, Boulet JR, Cooper JB, Petrusa ER, Baker KH, et al. Simulation-based assessment to reliably identify key resident performance attributes. *Anesthesiology*. 2018;128:821–31, <http://dx.doi.org/10.1097/ALN.0000000000002091>.

30. Lockyer J, Carraccio C, Chan MK, Hart D, Smee S, Touchie C, et al. Core principles of assessment in competency-based medical education. *Med Teach*. 2017;39:609–16, <http://dx.doi.org/10.1080/0142159X.2017.1315082>.
31. Castellanos-Ortega A, Broch MJ, Barrios M, Fuentes-Dura MC, Sancerni-Beitia MD, Vicent C, et al. Acceptance and validity of the methods used to implement a competency based medical education programme in an Intensive Care Department of a teaching referral center. *Med Intensiva*. 2021;45:411–20, <http://dx.doi.org/10.1016/j.medine.2019.12.011>.
32. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ*. 2011;45:1181–9, <http://dx.doi.org/10.1111/j.1365-2923.2011.04075.x>.
33. Khanduja PK, Bould MD, Naik VN, Hladkiewicz E, Boet S. The role of simulation in continuing medical education for acute care physicians: a systematic review. *Crit Care Med*. 2015;43:186–93, <http://dx.doi.org/10.1097/CCM.0000000000000672>.
34. Soffler MI, Claar DD, McSparron JI, Ricotta DN, Hayes MM. Raising the stakes: assessing competency with simulation in pulmonary and critical care medicine. *Ann Am Thorac Soc*. 2018;15:1024–6, <http://dx.doi.org/10.1513/AnnalsATS.201802-120PS>.
35. Murray DJ, Boulet JR, Boyle WA, Beyatte MB, Woodhouse J. Competence in decision making: setting performance standards for critical care. *Anesth Analg*. 2021;133:142–50, <http://dx.doi.org/10.1213/ANE.0000000000005053>.
36. Seam N, Lee AJ, Vennero M, Emlet L. Simulation training in the ICU. *Chest*. 2019;156:1223–33, <http://dx.doi.org/10.1016/j.chest.2019.07.011>.
37. Krage R, Erwtaman M. State-of-the-art usage of simulation in anesthesia: skills and teamwork. *Curr Opin Anaesthesiol*. 2015;28:727–34, <http://dx.doi.org/10.1097/ACO.0000000000000257>.
38. Weersink K, Hall AK, Rich J, Szulewski A, Dagnone JD. Simulation versus real-world performance: a direct comparison of emergency medicine resident resuscitation entrustment scoring. *Adv Simul (Lond)*. 2019;4:9, <http://dx.doi.org/10.1186/s41077-019-0099-4>.
39. Isaak RS, Chen F, Martinelli SM, Arora H, Zvara DA, Hobbs G, Stiegler MP. Validity of simulation-based assessment for accreditation council for graduate medical education milestone achievement. *Simul Healthc*. 2018;13:201–10, <http://dx.doi.org/10.1097/SIH.0000000000000285>.
40. Boet S, Bould MD, Fung L, Qosa H, Perrier L, Tavares W, Reeves S, Tricco AC. Transfer of learning and patient outcome in simulated crisis resource management: a systematic review. *Can J Anaesth*. 2014;61:571–82, <http://dx.doi.org/10.1007/s12630-014-0143-8>.