Oral Examination versus Simulation-Based Assessment in Assessing Patient Care Competencies for Emergency Medicine Residents

Wail Bamadhaf

Department of Emergency Medicine, Rashid Hospital, Dubai Health Authority, Dubai, UAE

Keywords
Miller’s pyramid of assessment · Oral examination · Simulation-based assessment · High-stakes assessment · Medical education

Abstract
Background: Assessment is essential for any accreditation process in the medical field. If a candidate passes a high-stakes assessment, they can work independently. While oral examinations are common, given the complexity of clinical competencies, such an approach may not be the most effective assessment method. A form of performance-based assessment, such as a simulation, may be beneficial in this context. Objectives: This study aims to determine whether the results of oral examinations match those of simulation-based assessments when both modalities are used to evaluate residents’ performance in scenarios featuring similar content. It also seeks to determine whether oral examinations under- or overestimate residents’ competencies concerning patient care when compared to their simulation performance. Methods: This is a cross-sectional, single-centre study. Emergency medicine residents underwent an oral examination and completed a simulation-based assessment. Standardized scenarios were used to assess the residents’ emergency medicine competencies. A global rating scale was used to rate participants’ performance in each assessment modality. Results: There was a moderate positive correlation between oral examination and simulation-based assessment results ($r = 0.699$, $p < 0.05$, $n = 28$). A paired t test indicates that the oral examination overestimates residents’ competency compared to the simulation-based assessment; the mean difference is 0.26 (confidence interval: 0.041–0.493). Conclusions: Emergency medicine residents whose knowledge was assessed at the “know-how” level of Miller’s pyramid in the oral examination were not necessarily able to move up to the level of “show-how” by demonstrating the ability to apply their knowledge in the simulation-based assessment. The findings of this study confirm that simulation-based assessments should be an essential aspect of high-stakes examinations intended to determine residents’ different clinical competencies.

Introduction
The accreditation process to obtain a specialist degree in any medical field, including emergency medicine, is challenging. After completing a residency programme, a
Simulation-Based Assessment in Emergency Medicine

resident must pass written and practical exams to obtain a specialization certificate. This assessment is considered high-stakes because a candidate who passes the test will be qualified to work independently. The emergency medicine board certification exam in the Middle East relies on written and practical assessments to determine residents’ competencies in patient care. The written assessment focuses on factual knowledge, whereas clinical skills are usually assessed by means of a practical assessment. The practical assessment may consist of both an objective structured clinical examination and an oral examination. The oral examination is a face-to-face assessment with two examiners who discuss various medical cases with the resident [1]. The resident cannot see the patient and can only listen to descriptions of physical examinations and view paper copies of the patient’s lab results and diagnostic imaging [1]. The cases used in the oral examination primarily feature acute life-threatening conditions that emergency medicine specialists are likely to encounter during their future clinical work. The examiners cannot assess residents managing these cases in actual clinical practice because doing so would pose a threat to patient safety.

The oral examination is intended to ensure that emergency medicine residents have acquired the patient care competencies needed to ensure that patients are provided with optimal treatment. This assessment modality assesses a resident’s competency at the “knows how” level in Miller’s pyramid, but it cannot determine their competency at the “shows how” level. In other words, the resident who “knows how” in the oral examination may not necessarily be able to “show how” in a simulation-based assessment [2]. Due to the complexity of clinical competencies, more sophisticated methods are needed to assess them; one such method is simulation, a form of performance-based assessment.

Given the importance of the oral examination, its lengthy history in emergency medicine, and the importance of simulation in medical education, particularly its validity and reliability for clinical competency assessments [3–5], it is worth comparing the effectiveness of these two means of assessments. This study compares the results obtained when using the traditional oral examination and a simulation-based assessment to determine residents’ competencies. This study hypothesizes that the oral examination overestimates residents’ patient care competence compared to the simulation-based assessment. The research presented in this work is crucial, as there is no study in the literature comparing the traditional oral examination for emergency medicine residents with a simulation-based assessment using a global rating scale in the Middle East. This research may encourage certification bodies to adopt a simulation-based assessment as a part of the certification process for emergency medicine residents.

Materials and Methods

Objectives

This study was designed to achieve the following objectives:

1. To determine whether the results of the traditional oral examination match residents’ performance in a simulation-based assessment using scenarios featuring similar content.
2. To determine whether the oral examination under- or overestimates residents’ patient care competency compared to their simulation-based performance.

Participants

After the author obtained approval from the Student Research Evaluation Committee and the Scientific Research Ethics Committee at the Dubai Health Authority, this study began. The study included third-, fourth-, and fifth-year emergency medicine residents working at the Rashid Hospital Trauma Centre who agreed to participate. Residents who had completed their training years and were awaiting an opportunity to repeat the board certification exam were excluded. Thirty-one residents fulfilled the inclusion criteria. In addition to providing written informed consent, the participants also signed a written confidentiality agreement to ensure that the content of the clinical scenarios would not be disseminated prior to the end of the study. Five senior emergency medicine residency program faculty members participated in the study by serving as a panel of examiners. The raters attended an educational workshop that outlined performance protocols applied in this study.

Study Design

A cross-sectional, single-centre study was conducted in the simulation rooms of the Rashid Hospital in July 2021. Each batch of participants was quarantined in the academic day hall on the assessment day and assessed based on one clinical scenario; each candidate completed one oral exam and performed one simulation-based assessment. The sessions lasted 30 min. In each session, two candidates were assessed. One of the candidates underwent the oral examination first, while the other carried out the simulation-based assessment (15 min were allocated for each encounter). Once the residents had completed each encounter, they immediately switched between assessment modalities. Participants were not informed that the content of the oral examination and simulation scenarios were the same; moreover, the opening scenarios differed between the two assessment modalities to avoid pattern recognition.

Three standardized patient management scenarios were developed. The cases were hyperkalaemia, β-blocker overdose, and adrenal insufficiency crisis. A special focus was placed on developing scenarios with simulation features. The key elements of the simulation scenarios resembled the material covered in the oral examination, thus allowing for meaningful comparisons between the two assessment modalities. The oral script took the form of the
Scoring System and Scoring Process
Differences in the participants’ performance were assessed using the global rating scale on each assessment modality. Each batch was randomly assigned a pair of examiners. The scale consisted of nine components that assess the residents’ performance in taking patient histories, performing clinical examinations, formulating differential diagnoses, ordering appropriate laboratory tests, solving illness-related problems, managing therapeutic treatment, reaching an accurate final diagnosis, requesting appropriate consultations service, and demonstrating communication skills. Each component is rated on a five-point scale where 1 indicates significant omissions and inappropriate or irrational practices and 5 indicates a comprehensive or exemplary approach. A narrative record section at the end of the scale allows the examiner to record any critical actions made during patient care, including any major errors. The faculty members who participated in the study were trained to use the scoring system in the educational workshop. They were informed that a performance score of 3 would be the minimum acceptable level for an independent practitioner. Furthermore, they were instructed to observe the quality of the cardiopulmonary resuscitation performed in the simulation-based assessment and record any comments on the narrative portion of the scoring sheet. No scores or feedback was provided to the candidates during the evaluation. The nine components were summed to obtain an overall average score indicating the candidates’ performance for each modality (minimum score 1, maximum score 5). The two scores assigned to the participants by the different examiners in the same modality were averaged to produce a final score. After completing the study, candidates were individually debriefed by the examiners on their performance.

Statistical Analysis
Statistical analysis was performed using SPSS 25. Numerical data were used to achieve the objectives of the study. A normality analysis was conducted, and a normal Q-Q plot was developed. It was found that the data followed a normal distribution, and parametric tests could thus be employed. Participants’ scores were correlated between assessment modalities using the Pearson correlation coefficients and inter-class correlation coefficient (ICC). A paired \( t \) test analysis was also conducted to evaluate the difference in performance between the two modalities; a \( p \) value of less than 0.05 was considered statistically significant.

Results
Twenty-eight of the 31 eligible residents participated in the study. Three residents, each of whom was in a different year, did not attend the assessment days due to illness. A correlation analysis was conducted to test the correlation between the oral examination results and the simulation performance results. Furthermore, a paired \( t \) test analysis was conducted to evaluate the difference in performance between the two modalities.

Correlation between Assessment Modalities
A Pearson correlation was conducted to compare the performance of the two modalities used in the study, namely, the oral examination and the simulation-based examination. The analysis indicated the presence of a positive correlation between the oral examination and the simulation-based examination in terms of performance results. The correlation \((r = 0.699^{* *}, p < 0.05, n = 28)\) indicates that the residents’ performance levels as measured by the oral modality correlate moderately with that measured through the simulation modality. Therefore, it can be concluded that a moderate positive correlation exists between residents’ performance in the oral examination and that in the simulation-based assessments.

In addition, to achieve the first research objective, it was necessary to evaluate the agreement of performance levels between the two modalities. For this purpose, the ICC was calculated. The results indicated that there was a moderate degree of agreement between the performance measurements of the oral and the simulated examinations. The average measure ICC was 0.783, showing high significance with a 95% confidence interval (CI: 0.345–0.820).

Paired \( t \) Test
A paired \( t \) test was conducted to determine whether the oral examination underestimated or overestimated the residents’ patient care competency compared to their simulation-based performance. In addition, the test was also conducted to determine whether there was a considerable difference between the two groups; the difference was determined using the means of both groups [6].

As indicated in Table 1, the mean score obtained for the oral examination was 3.7177, while that for the simulated assessment was 3.4505. As shown in Table 2, there was a difference between the oral and the simulation examinations (mean difference = 0.26714, \( t = 2.425 \)). This indicates a significant difference between the oral and simulation-based performance. A correlation analysis was conducted to test the correlation between the oral and simulation examination results and the simulation-based performance results. Furthermore, a paired \( t \) test analysis was conducted to evaluate the difference in performance between the two modalities.

Table 1. Paired sample statistics

| Modality            | Mean  | N  | Std. deviation | Std. error mean |
|---------------------|-------|----|----------------|-----------------|
| Oral examination    | 3.7177| 28 | 0.63617        | 0.12023         |
| Simulated examination| 3.4505| 28 | 0.80904        | 0.15289         |
finding indicates that the oral examination overestimated the residents’ patient care competency compared to their simulation-based performance.

**Discussion**

The study’s first objective was to determine whether the residents’ oral examination results matched their performance results in the simulation-based assessment when both featured similar scenarios. Pearson correlation coefficients and ICCs were calculated. The Pearson correlation coefficients showed a moderate positive correlation ($r = 0.699$, $p < 0.05$, $n = 28$). These results align with the finding of Savoldelli et al. [2] and Schwid et al. [7] that residents’ performance outcomes in oral examinations correlate moderately with those in simulated-based examinations. Moreover, the average measure ICC was 0.783, showing high significance with a 95% CI (0.345–0.820). This result indicates that there is moderate agreement between these two modalities. I cannot find any evidence in the literature indicating no correlation between these two modalities. The raw data indicated that residents who scored low on the oral examination also scored low in the simulation-based assessment and vice versa. This observation affirms the correlation between the two assessment modalities. In my opinion, if a resident has no theoretical knowledge and understanding of a topic, they will be unable to practically address any cases involving that topic.

Simulation-based learning is a well-established teaching technique in the medical field [8]. This study focused on using simulation as an assessment modality in high-stakes examinations. Although a similar rating scale was used for both assessment modalities, the scoring process differed. Oral examination scoring indicates a resident’s theoretical knowledge of an illness and ability to reason logically, effectively verbalize treatment plans, and articulate opinions plainly and concisely. In contrast, in a simulation-based assessment, examiners primarily score the logical sequencing of treatment actions and demonstration of technical and communication skills. An oral examination helps examiners assess residents’ theoretical knowledge and the application thereof in practice; however, examiners cannot determine how residents would use this information and whether or not they would be able to effectively and independently manage a patient in practice.

The study’s second objective was to determine whether the oral examination under- or overestimates residents’ ability to provide patient care when compared to the simulation-based assessment. It was expected that the oral examination would yield higher performance results than the simulated-based assessment, and the results confirmed this. The mean difference of 0.26 (CI: 0.041–0.493) suggests that the oral examination overestimates residents’ patient care competency. This result is significant and in line with the central hypothesis of this study. In addition, this result is comparable to that of Schwid et al. [7], who found that an oral examination overestimated anaesthesia residents’ competencies compared to a simulation-based assessment. However, they did not perform a paired $t$ test for the two modalities. They noted that residents who passed the oral test made numerous treatment errors in the simulation assessment. I believe that these errors were due to the fact that the oral examination score reflects a resident’s ability to explain a problem and implement a management plan to resolve it, not whether they can perform well in practice. In contrast, in the simulation-based assessment, practical performance is assessed more critically. This was evident in the examiners’ comments in the narrative part of the scoring sheet in the current study. Residents failed to perform high-quality cardiopulmonary resuscitation in the simulation-based assessment even though they did emphasize the importance of this skill in the oral examination. Savoldelli et al. [2] and Schwid et al. [7] obtained similar findings. The authors found that residents scored higher in an oral assessment than in a simulation-based assessment because the residents could more easily explain how a medical

| Paired difference | T   | df | Sig. (2-tailed) |
|-------------------|-----|----|----------------|
| Oral examination – simulated examination | 0.26714 | 0.58292 | 0.11016 | 0.04111 | 0.49318 | 2.425 | 27 | 0.022 |
condition should be treated than practically treat that condition in a simulation.

Using simulation-based assessments for high-stakes examinations remains a controversial topic in the Middle East. Due to the complexity of medical competencies, no gold-standard assessment modality can capture the full scope of residents’ clinical competencies [9]. Despite the well-documented limitations of the oral examination, it remains the sole assessment modality for the practical section of the board certification examination. It is unlikely to disappear in the future. The board certification body should consider the use of different modalities to cover the different levels of Miller’s assessment pyramid. A multimodal assessment approach could be used to assess various clinical competencies of candidates.

Limitations of the Study
This study is subject to two main limitations, which are due to the participants and the provided scenarios. The first limitation is the small sample size. There were only 28 resident participants, all of whom were from one academic training site. This limitation may have affected the study results, as all the residents follow the same academic curriculum in their residency program. Although they practice simulations regularly during their training years, these residents have more training in oral examinations in preparation for the board certification examination, which may have affected their performance during the simulation-based assessment.

The second limitation is that residents’ performance was assessed in only three scenarios. The small number of scenarios limits the generalizability of the results; using multiple scenarios would have increased generalizability. Future studies should include a larger sample of residents across numerous centres and a broader range of scenarios to increase the generalizability of their results.

Conclusion
The results of this study confirmed that compared with simulation-based assessment, oral examination overestimates residents’ competencies, implying that competency varies depending on assessment modality. Emergency medicine residents whose knowledge was assessed at the level of “know-how” in the oral examination were not necessarily able to move up to the next level in Miller’s pyramid, “show how,” in the simulation-based assessment. For this reason, simulation should become an essential element of high-stakes examinations intended to determine residents’ various clinical competencies.

Acknowledgements
I am thankful to Dr. Sara Kazim for all the support she offered during my study. Also, thanks to Dr. Firas AlNajjar, whose knowledge helped me during this research. Finally, I thank the Emergency Medicine Residency Program faculty at the Rashid hospital trauma centre, Dr. Aiman Magzooob, Dr. Chafika Lasfer, Dr. ara S.M Abumuaileq, Dr. Maryam Al Ali, Dr. Malik Zakaullah, and Dr. Lubna Saffarini.

Statement of Ethics
The study was approved by the Dubai Scientific Research Ethics Committee in Dubai Health Authority (DSREC-SR-06/2020_01). Written Informed consent was obtained from all participants of the study. The participants have consented to the submission of the study for publication. This research complies with the guidelines for human studies in accordance with the World Medical Association Declaration of Helsinki.

Conflict of Interest Statement
The author has no conflicts of interest to declare.

Funding Sources
There has been no financial support for this work that could have influenced its outcome.

Data Availability Statement
The data that support the findings of this study are not publicly available due to privacy concerns regarding the research participants but are available from W.B. upon reasonable request.

References
1 Solomon DJ, Reinhart MA, Bridgham RG, Munger BS, Starnaman S. An assessment of an oral examination format for evaluating clinical competence in emergency medicine. Acad Med. 1990;65(9):S43–4.
2 Savoldelli GL, Naik VN, Joo HS, Houston P, Graham M, Yee B, et al. Evaluation of patient simulator performance as an adjunct to the oral examination for senior anesthesia residents. Anesthesiology. 2006;104(3):475–81.
3 Hart D, Bond W, Siegelman JN, Miller D, Cassara M, Barker L, et al. Simulation for assessment of milestones in emergency medicine residents. *Acad Emerg Med*. 2018;25(2):205–20.

4 Bond WF, Lammers RL, Spillane LL, Smith-Coggins R, Fernandez R, Reznek MA, et al. The use of simulation in emergency medicine: a research agenda. *Acad Emerg Med*. 2007;14(4):353–63.

5 Girzadas DV Jr, Clay L, Caris J, Rzechula K, Harwood R. High fidelity simulation can discriminate between novice and experienced residents when assessing competency in patient care. *Med Teach*. 2007;29(5):472–6.

6 Kim TK. T test as a parametric statistic. *Korean J Anesthesiol*. 2015;68(6):540.

7 Schwid HA, Rook GA, Carline J, Steadman R, Murray W, Olympio M, et al. Evaluation of anesthesia residents using mannequin-based simulation: a multiinstitutional study. *Anesthesiology*. 2002;97(6):1434–44.

8 Jahanshir A, Bahrinei M, Banaie M, Jallili M, Hariri S, Rasooli F, et al. Implementation a medical simulation curriculum in emergency medicine residency program. *Acta Med Iran*. 2017;55(8):521–4.

9 Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945–9.