

Education

Development and evaluation of a simulation-based resuscitation scenario assessment tool for emergency medicine residents

Andrew Koch Hall, MD*; William Pickett, MSc, PhD*; Jeffrey Damon Dagnone, MD, MMed*

ABSTRACT

Objective: We sought to develop and validate a three-station simulation-based Objective Structured Clinical Examination (OSCE) tool to assess emergency medicine resident competence in resuscitation scenarios.

Methods: An expert panel of emergency physicians developed three scenarios for use with high-fidelity mannequins. For each scenario, a corresponding assessment tool was developed with an essential actions (EA) checklist and a global assessment score (GAS). The scenarios were (1) unstable ventricular tachycardia, (2) respiratory failure, and (3) ST elevation myocardial infarction. Emergency medicine residents were videotaped completing the OSCE, and three clinician experts independently evaluated the videotapes using the assessment tool.

Results: Twenty-one residents completed the OSCE (nine residents in the College of Family Physicians of Canada–Emergency Medicine [CCFP-EM] program, six junior residents in the Fellow of the Royal College of Physicians of Canada–Emergency Medicine [FRCP-EM] program, six senior residents in the FRCP-EM). Interrater reliability for the EA scores was good but varied between scenarios (Spearman $\rho = [1] 0.68, [2] 0.81, [3] 0.41$). Interrater reliability for the GAS was also good, with less variability ($\rho = [1] 0.64, [2] 0.56, [3] 0.62$). When comparing GAS scores, senior FRCP residents outperformed CCFP-EM residents in all scenarios and junior residents in two of three scenarios ($p < 0.001$ to 0.01). Based on EA scores, senior FRCP residents outperformed CCFP-EM residents, but junior residents outperformed senior FRCP residents in scenario 1 and CCFP-EM residents in all scenarios ($p = 0.006$ to 0.04).

Conclusion: This study outlines the creation of a high-fidelity simulation assessment tool for trainees in emergency medicine. A single-point GAS demonstrated stronger relational validity and more consistent reliability in comparison with an EA checklist. This preliminary work will provide a foundation for ongoing future development of simulation-based assessment tools.

RÉSUMÉ

Objectif: Nous avons travaillé à concevoir et à valider un outil d'examen clinique objectif structuré (ECOS), reposant sur la simulation et comptant trois postes, pour évaluer la compétence des résidents en médecine d'urgence, dans différents scénarios de réanimation.

Méthode: Un groupe d'experts composé d'urgentologues a conçu trois scénarios dans lesquels il y avait des mannequins d'apparence très réaliste. Pour chaque scénario, les experts ont élaboré un outil correspondant d'évaluation, qui comprenait une liste de vérification d'actions essentielles (AE) et un résultat d'évaluation globale (REG). Les scénarios portaient sur 1) la tachycardie ventriculaire instable, 2) l'insuffisance respiratoire, et 3) l'infarctus du myocarde avec sus-décalage du segment ST. Les résidents en médecine d'urgence ont été filmés sur bande vidéo au cours de l'ECOS, et trois experts cliniciens ont évalué, chacun de leur côté, les vidéos à l'aide des outils d'évaluation.

Résultats: Vingt et un résidents ont fait l'ECOS (neuf résidents inscrits au programme de médecine d'urgence du Collège des médecins de famille du Canada [CMFC-MU], six résidents juniors inscrits au programme de médecine d'urgence des Associés du Collège royal des médecins et chirurgiens du Canada [FRCP-MU], et six résidents seniors inscrits au FRCP-MU). La fiabilité interévaluateurs en ce qui concerne les résultats des AE était bonne, mais elle variait entre les scénarios (coefficient de corrélation des rangs de Spearman [CCRS] = [1] 0.68, [2] 0.81, [3] 0.41). La fiabilité interévaluateurs en ce qui concerne le REG était également bonne, mais elle avait une variabilité moindre (CCRS = [1] 0.64, [2] 0.56, [3] 0.62). Dans les comparaisons des REG, les résidents seniors du FRCP ont obtenu de meilleurs résultats que les résidents du CMFC dans tous les scénarios et les résidents juniors dans deux scénarios sur trois ($p, 0.001$ à 0.01). D'après les résultats des AE, les résidents seniors du FRCP ont surclassé les résidents du CMFC, mais les résidents juniors ont surclassé les résidents seniors du FRCP dans le premier scénario et les résidents du CMFC dans tous les scénarios ($p = 0.006$ à 0.04).

From the *Department of Emergency Medicine, Queen's University, Kingston, ON.

Correspondence to: Dr. Andrew Koch Hall, Department of Emergency Medicine, Kingston General Hospital, Empire 3, 76 Stuart Street, Kingston, ON K7L 2V7.

This article has been peer reviewed.

© Canadian Association of Emergency Physicians

CJEM 2012;14(3):139-146

DOI 10.2310/8000.2012.110385

DECKER



CJEM • JCMU

2012;14(3) 139

Conclusion: L'étude décrit, dans ses grandes lignes, la conception d'un outil d'évaluation reposant sur des simulations réalistes pour les stagiaires en médecine d'urgence. Un REG à un seul point a révélé une validité relationnelle plus forte et une fiabilité plus constante que la liste de vérification des AE.

Ce travail préliminaire servira de base à l'élaboration future et continue d'outils d'évaluation fondés sur la simulation.

Keywords: assessment, emergency medicine, medical education, resident, resuscitation, simulation

Assessment of clinical competence in postgraduate medical education is moving away from knowledge-based examination and toward outcome-based assessment.¹ As noted by Miller, through medical training, there is a progression from “knows” to “knows how,” “shows how,” and “does.”² Knowledge acts as a foundation for clinical reasoning and decision making, which, in turn, supports successful clinical performance. The direct evaluation of performance, through simulation-based assessment, provides a unique opportunity for simultaneous evaluation of knowledge, clinical reasoning, and teamwork.³ Emergency medicine (EM) training programs would likely benefit from competency-based assessment tools to evaluate how trainees perform in acute care scenarios.^{4,5} High-fidelity simulation-based assessment tools with standardized scenarios have the potential to achieve this goal.^{6,7}

High-fidelity simulations use computer-controlled mannequins supported by mechanized movements, cues, and simulated vital signs to emulate real patient encounters. Simulation-based education provides the opportunity to encourage deliberate practice of standardized learning experiences even with rare disease states, while simultaneously improving the imperative skills of teamwork and communication.^{8,9} Over the past decade, there has been a dramatic increase in the use of high-fidelity simulation in EM residency programs. Okuda and colleagues reported that 85% of American EM residency programs surveyed in 2008 used mannequin-based simulators in their curriculum.¹⁰ In some cases, simulation has been directly incorporated into postgraduate EM curricula, not as an adjunct but as the primary education strategy for topics that have been deemed to be best taught with a simulation format.¹¹

The Accreditation Council for Graduate Medical Education (ACGME) specifically recommends the use of simulation to assess medical diagnostic, treatment, and procedural skills.¹² Despite the integration of high-fidelity mannequin-based simulation into postgraduate medical curricula, there has been limited research in the development of validated evaluation tools for simulation-based assessment.⁷ Recently, studies in

pediatrics and anesthesia have demonstrated the ability of high-fidelity simulation assessment tools using case scenarios to discriminate between trainees of different levels with good interrater reliability.^{13–17} A simulation-based Objective Structured Clinical Examination (OSCE) has been incorporated into the Israeli National Board Examination in Anaesthesia.¹⁸ This is the first reported case of this type of tool being used for high-stakes examination. Girzadas and colleagues developed an assessment tool using high-fidelity simulation that discriminated between novice and experienced EM residents in the management of a case of anaphylaxis requiring a surgical airway.¹⁹

The postgraduate EM training program at Queen's University has integrated high-fidelity simulation through weekly junior and senior resuscitation rounds and monthly core curriculum sessions. We previously reported the pilot evaluation of an assessment tool for a standardized ventricular fibrillation cardiac arrest scenario.²⁰ The purpose of this study was to develop and validate a three-station simulation-based OSCE tool to assess EM resident competency in resuscitation scenarios.

METHODS

Study design

A cross-sectional study with a validation component was employed. We developed three standardized emergency department resuscitation scenarios, each with a corresponding assessment tool. They were administered in a videotaped OSCE format to 22 EM residents over 2 days in our simulation laboratory in August 2009. Residents were instructed to maintain confidentiality about the content of the OSCE between sessions. A written examination of similar content was administered to each participant immediately following the OSCE. The videotaped performances were then scored by three attending EM faculty using the assessment tools. We hypothesized that level of training would correlate positively with performance on the OSCE on both the global

assessment score (GAS) and essential action (EA) scores. Additionally, we hypothesized that there would be a linear correlation between EA scores and the GAS.

Study participants

Twenty-two residents were enrolled: 12 from the Fellow of the Royal College of Physicians of Canada–Emergency Medicine (FRCPEM) program (six junior [postgraduate year (PGY) 1–2] residents and six senior [PGY 3–5] residents) and 10 from the College of Family Physicians of Canada–Emergency Medicine certificant (CCFPEM) program. Participation was voluntary, and all residents provided written informed consent to participate. This study was reviewed and approved by the Queen’s University Ethics Review Board. All residents had participated in at least four simulation-based resuscitation sessions in our simulation laboratory within the 2 months prior to the examination.

Scenario and assessment tool development

An expert panel of EM faculty, all with training in high-fidelity simulation-based instruction, developed three scenarios to be used as OSCE stations. The chosen scenarios were (1) unstable ventricular tachycardia, (2) respiratory failure, and (3) ST elevation myocardial infarction (STEMI). All scenarios were based on content previously covered in simulation-based teaching sessions in the 2 months prior to the study, which included Advanced Cardiac Life Support (ACLS) guidelines and advanced airway management. Each station was designed to last 5 to 10 minutes. The scenarios were resuscitation focused and designed to discriminate between residents of different levels of training by eliciting observable behaviours that traditionally are poorly assessed with written examinations. Each scenario included scripted roles and clear instructions for a nurse, the mannequin technician, and, for scenario 2, a respiratory technician.

OSCE 2009				
RAPID SEQUENCE INTUBATION				
	NOT DONE 0	DONE 1	DONE WELL 2	UNABLE TO OBSERVE
PRE-OXYGENATION 100% O2 by BVM		NRB	BVM	
Optimization of PATIENT POSITIONING <small>(supine, head of bed, pillow, bed height)</small>		min 1/4	min 3/4	
	NOT DONE 0	DONE 1	DONE WELL 2	UNABLE TO OBSERVE
INDUCTION AGENT choice		Ketamine	Etomid/Prop	
INDUCTION AGENT correct dose <small>(KETAMINE 1-2mg/kg ETOMIDATE 0.3mg/kg PROPOFOL 1-2mg/kg)</small>				
PARALYTIC AGENT given		SGC, RSC		
PARALYTIC AGENT correct dose <small>(SGC 1.5-2mg/kg ROCURONIUM 1mg/kg)</small>				
	NOT DONE 0	DONE 1	DONE WELL 2	UNABLE TO OBSERVE
LARYNGOSCOPY timed appropriately				
LARYNGOSCOPY performed safely				
ETT placement performed safely				
ETT confirmation with ETCO2			Delayed	
ETT confirmation with AUSCULTATION				
pCXR ordered				
POST-intubation sedation (+/- paralysis)				
OG / NG placement				
TOTAL SCORE				
GLOBAL ASSESSMENT				
	1	2	3	4
	Fail	Marginal Pass	Satisfactory	Above Average
				5 Excellent

Figure 1. Scenario 2 assessment tool. OSCE = Objective Structured Clinical Examination.

Assessment tools followed a standardized format based on our previous pilot work²⁰ and were composed of two separate components: EAs and a single GAS. An example assessment tool is provided as Figure 1. Each EA checklist contained 14 to 21 items, which used either dichotomous response categories (action completed or not completed) or a three-level weighted score (completed well, completed poorly, or not completed). The three-level scores were used for actions deemed more critical for successful management of the scenario. Each checklist item was derived and reviewed using a modified Delphi technique by a panel of emergency and intensive care physicians from the Kingston Resuscitation Institute who were not directly involved in the study. The GAS was assigned on a Likert scale ranging from 1 (fail) to 5 (excellent). All of the scenarios and assessment tools were piloted and revised in the simulation laboratory through practice sessions with the involved EM faculty and support staff.

A formative multiple-choice and short-answer examination was administered to all residents at the conclusion of the OSCE to assess their baseline medical knowledge of the scenario subject matter. This was used for feedback purposes to aid learning, not to assign a specific score contributing to the candidate's evaluation within the residency program. The marks assigned on the written questions, which involved material related to the OSCE scenarios (tachycardias, airway medications, STEMI), were collected and compared to simulator performance.

OSCE administration and evaluation

The three-station OSCE was administered to the residents over 2 separate days. Each resident was

scheduled in a 20-minute block, and the scenarios were performed in order. The resident was provided with standardized verbal instructions prior to each scenario and given 1 minute to read a brief introduction to the scenario attached to the door of the room. The residents' performance was videotaped from a fixed camera angle, which allowed simultaneous visualization of the examinee, the mannequin, and the cardiac monitor. Members of the research team observed each session from behind one-way glass. We used a METI Emergency Care Simulator (Medical Education Technologies Inc, Sarasota, FL) high-fidelity simulation mannequin in a dedicated simulation laboratory designed to recreate the physical environment of an emergency department resuscitation bay. All candidates were debriefed for 5 minutes about their performance and then given 30 minutes to complete the written examination in a separate room.

The videotapes were reviewed by three independent content experts. All reviewers were attending EM faculty with 5 years or more of experience and specific training in simulation-based education. To standardize evaluations, the reviewers received a 2-hour training session on the use of the assessment tools with feedback from one of the investigators (J.D.D.). All reviewers were asked to rate a sample of training videotape as poor, adequate, and excellent performances for each scenario. These were then reviewed individually with the investigator until predefined standardized scores were agreed upon.

Data analysis

Data analysis was performed using SPSS version 17.0 (SPSS Inc, Chicago, IL). Items not scored by the

	Scenario 1 Unstable VT	Scenario 2 RSI	Scenario 3 STEMI
Essential action score			
Mean score, <i>n</i> (%)	17.4/24 (72.5)	12.3/21 (58.6)	17.0/30 (56.7)
Range	11.3–21.3	5.7–17.3	14.3–21.0
IRR (r_s)	0.68	0.81	0.41
Global assessment score			
Mean score, <i>n</i> (%)	3.1/5 (62)	3.2/5 (64)	3.2/5 (64)
Range	1.7–4.7	1.3–5.0	1.0–4.7
IRR (r_s)	0.64	0.56	0.62

IRR = interrater reliability; r_s = mean Spearman correlation coefficient; RSI = rapid-sequence intubation; STEMI = ST elevation myocardial infarction; VT = ventricular tachycardia.
Mean, minimum, and maximum scores for essential action scores and the global assessment score are based on summary values across all raters.

reviewers were given a score of 0. Interrater reliability for EA scores and the GAS for each of the three scenarios were estimated using the Spearman rank correlation coefficient. To analyze the discriminatory capabilities of the assessment tool, the GAS and EA scores were compared by seniority (junior FRCP [PGY 1–2], CCFP-EM [PGY 3], and senior FRCP [PGY 3–5]) using analysis of variance (ANOVA). Finally, we calculated the degree of correlation between the EA scores and the GAS with results from the written examination using the Spearman coefficient.

RESULTS

Twenty-two of the 26 Queen's University residents registered in the two EM programs were enrolled. Two residents were outside the province due to fellowship training, one was on maternity leave, and one was a study investigator. There was a technical error in the videotaping file for one of the residents, which left 21 residents in the final analysis (6 junior FRCP-EM, 6 senior FRCP-EM, and 9 CCFP-EM residents).

Examinees' performance and the interrater reliabilities of the assessment tools for each of the three OSCE scenarios are shown in Table 1. The interrater reliabilities were fair to very good, ranging from 0.41 to 0.81. There was good correlation between EA scores and GAS in scenarios 1 ($r = 0.66$) and 2 ($r = 0.63$) but poor correlation in scenario 3 ($r = 0.38$). Individual trainee GAS and EA scores combined across all three scenarios are graphically provided in Figure 2.

A comparison of examinee performance by level of training is provided in Table 2, and examinee scores combined across all three scenarios are displayed in Figure 3. In two of three scenarios, senior FRCP residents outperformed all other residents in GAS and EA scores. In scenario 1, however, junior FRCP residents outperformed all residents in both GAS and EA scores. In scenarios 2 and 3, the CCFP-EM residents outperformed the junior FRCP residents in GAS, but the junior FRCP residents outperformed the CCFP-EM residents in EA scores. With respect to the written examination, there is a clear trend toward senior FRCP residents outperforming CCFP-EM residents, who outperformed junior FRCP residents in almost all scenarios. Correlation between performance on the relevant written examination and the OSCE performance was fair for EA scores ($r = 0.48$), and good for the GAS ($r = 0.60$).

DISCUSSION

Most postgraduate medical training programs in EM have integrated or are in the process of integrating high-fidelity mannequin-based simulation into their curricula. In this study, we outlined the development and evaluation of a multiple-station high-fidelity simulation-based OSCE for the objective evaluation of EM residents in standardized resuscitation scenarios.

In general, assessment is intended to quantify an underlying construct that is not easily amenable to direct measurement—in this case, competence in EM resuscitation scenarios. The assessment of validity

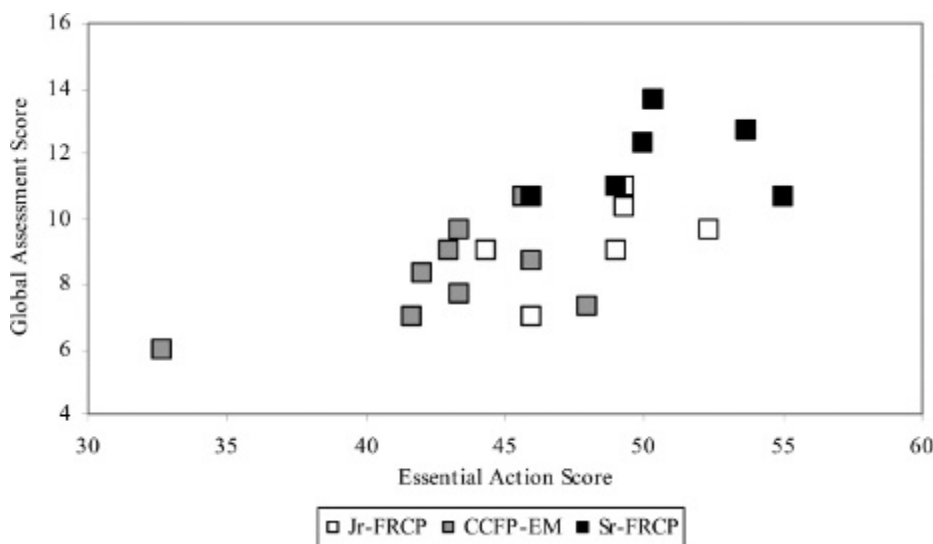


Figure 2. Individual resident essential action and global assessment scores combined across all scenarios by level of training. CCFP = certificant of the College of Family Physicians; EM = emergency medicine; FRCP = fellow of the Royal College of Physicians of Canada.

Table 2. Comparison of mean scores based on resident level of training by scenario

	Jr-FRCP (n = 6)	CCFP-EM (n = 9)	Sr-FRCP (n = 6)	p value
Scenario 1: Unstable VT				
Essential action score (/24)	19.2	15.9	17.8	0.03
Global assessment score (/5)	3.8	2.5	3.4	0.01
Relevant written examination (/16)	12.0	12.9	13.8	0.37
Scenario 2: RSI				
Essential action score (/21)	11.8	11.1	14.4	0.04
Global assessment score (/5)	2.9	3.0	4.0	0.003
Relevant written examination (/16)	10.7	12.4	14.5	0.001
Scenario 3: STEMI				
Essential action score (/30)	17.3	15.9	18.4	0.006
Global assessment score (/5)	2.6	2.9	4.3	< 0.001
Relevant written examination (/16)	8.4	8.4	11.0	0.05

CCFP = certificant of the College of Family Physicians; EM = emergency medicine; FRCP = fellow of the Royal College of Physicians and Surgeons of Canada; RSI = rapid-sequence intubation; STEMI = ST elevation myocardial infarction; VT = ventricular tachycardia.

involves multiple components, including content validity, internal structure validity, process validity, relational validity, and reliability.²¹ Content validity, or the extent to which included content is relevant to our defined construct, was ensured through the initial development by an expert panel of EM physicians and the subsequent review by an independent group with content expertise in resuscitative medicine. The content included was based on accepted principles of medical management, such as ACLS protocols and accepted standards in rapid-sequence intubation.²² The internal structure validity of the tool was strengthened by the selective weighting of actions in the EA checklist deemed to be more critical with a higher

potential score. Process validity, or the extent to which the cognitive and physical process required by the assessment represents the construct, is generally strong in high-fidelity mannequin-based simulation.²¹ We took a number of steps to ensure that other physical and procedural aspects of the room accurately reflected a typical resuscitation environment, including appropriate environment set-up and appearance, skill and demeanour of the assistants, and proper functioning of the mannequin.

Our findings of generally good discrimination between residents of different levels of training support the relational validity of the tool we developed. As expected, significant variability in performance was

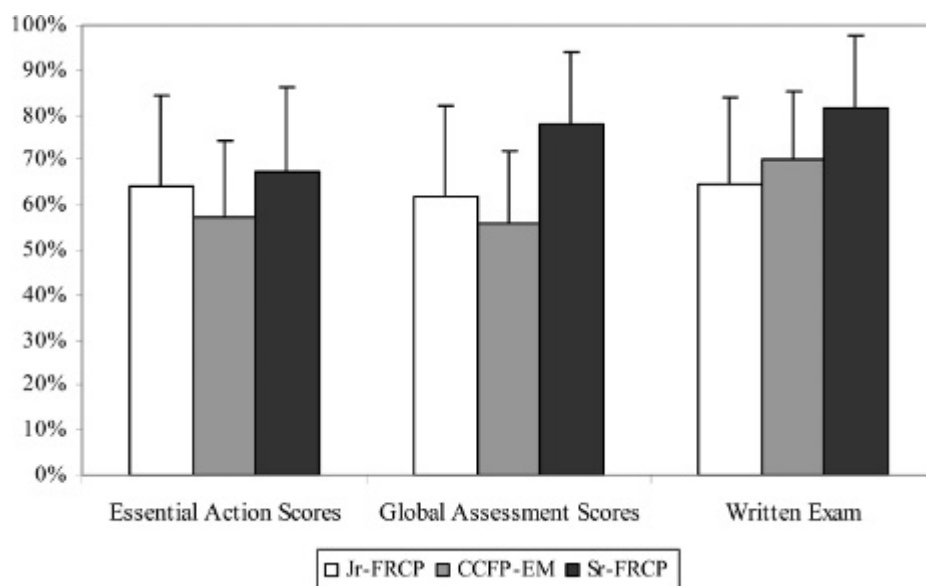


Figure 3. Mean resident essential action, global assessment scores, and written examination scores combined across all scenarios by level of training (+1 SE indicated by upper bar). CCFP = certificant of the College of Family Physicians; EM = emergency medicine; FRCP = fellow of the Royal College of Physicians of Canada.

seen within each level of training, but we did find a general trend toward higher scores with increasing level of training. With respect to the comparison of EA checklist scores, the junior FRCP residents (PGY 1–2) outperformed the CCFP-EM (PGY 3) residents and all residents in scenario 1. This was unexpected but may be explained by the timing of the examination and the background training of each group. At our institution, the junior FRCP residents conduct weekly resuscitation rounds in the simulation laboratory independent of senior residents (PGY 3–5) and core academic rounds. These start in the first week of residency and continue throughout the academic year. This provides the junior residents with significantly more experience in the simulation laboratory in the 2 months leading up to the OSCE when compared to the variable exposure of the CCFP-EM group, whose experience in high-fidelity simulations can be as little as four core training sessions. The probable effect of this was most obvious when observing the results of the ventricular tachycardia with cardioversion scenario, which the junior residents would have performed repeatedly, even weekly, in their resuscitation rounds immediately leading up to the OSCE. Although the senior residents had undergone similar weekly resuscitation rounds previously, it was likely the recent repetitive practice of the junior residents in this relatively straightforward resuscitation case that led to them scoring higher on scenario 1. Because the junior residents also scored higher than the others in the GAS in this scenario, their score is a direct reflection of their excellent performance rather than a flaw in the construct validity of the tool.

In general, we found that the GAS was able to accurately discriminate between residents of different levels of training, but there also seemed to be a linear correlation between the GAS and EA scores. Although previous work has demonstrated a good correlation between GAS and EA scores,²³ it is evident that each method of assessment carries specific drawbacks and benefits. EA checklists perform well when there are clearly delineated steps or components of performance that are well established, for example, involving procedural skills.⁵ The GAS is felt to be less precise than an EA checklist but may carry stronger validity in performance-based assessment.⁶ In the objective assessment of surgical residents, a GAS was shown to be more effective, with higher reliability and validity, than a checklist-based score.²⁴ GASs may capture aspects of

expert resuscitation that are difficult to delineate with a checklist, such as efficiency, decisiveness, and communication skills.⁶ This may suggest that a combination of the two scores or a hybrid scoring system would be most effective.¹ Our findings of a linear correlation between the GAS and EA scores and a superior discriminatory performance of the GAS support the future use of GAS alone or hybrid GAS with anchors.

The interrater reliability of GAS and EA scores for each scenario varied, but our results were comparable to those of other similar studies, with published mean reliability estimates of approximately 0.60 to 0.85.^{13–18} It is important to note that most of the other studies referenced are from anesthesia and pediatrics and use different assessment tools. We suspect that the poor reliability for scenario 3 arose because of errors in the development of EA checklist points as scenario 3 also had the lowest correlation between EA scores and GAS. EA checklist items are likely not specific enough to discriminate between poor, average, and excellent performances.

The limitations of our study design include a small sample size, incomplete blinding of raters, and the substantial effort and resources required to run and evaluate the simulations. We chose to focus only on EM residents at our institution, and our expert raters were from the same institution. Future studies may choose to use expert raters from a different institution to avoid bias. The estimated time required to evaluate all videotaped performances was 8 to 10 hours per evaluator, and the preparation and execution time for the examinations was over 30 hours and involved two physicians, one resident, one technician, and two nurses. It is clear that significant evaluator dedication and substantial resources are required to carry out simulation-based assessment. Our hope is that as our scenarios are validated and experience is gained, the process will become more streamlined and efficient.

Our study demonstrates the ability of high-fidelity simulation to be used as an assessment tool for EM resident competency. Our study findings demonstrate the great potential for the creation of valid assessment tools using high-fidelity simulation for use in high-stakes summative evaluations. We have highlighted the utility of a GAS and the difficulties associated with a checklist-based scoring system in the assessment of expertise in resuscitation. We plan to continue to develop, evaluate, and refine a collection of resuscitation scenarios to be used for high-fidelity simulation

training and assessment. This study and others of its kind can serve as the foundation for the further integration of high-fidelity simulation into summative and high-stakes assessment in the future. We expect that high-fidelity simulation will soon become an integral part of the summative assessment of residents in the form of both internal and across-center residency examinations and board certification examinations.

CONCLUSIONS

This study outlined the creation of a high-fidelity simulation assessment tool for trainees in EM. A single-point GAS demonstrated stronger relational validity and more consistent reliability in comparison with an EA checklist. This preliminary work will provide a foundation for ongoing future development of simulation-based assessment tools.

Competing interests: This study was supported by a Queen's Research Initiation Grant.

REFERENCES

- McLaughlin S, Fitch MT, Goyal DG, et al. Simulation in graduate medical education 2008: a review for emergency medicine. *Acad Emerg Med* 2008;15:1117-29, doi:[10.1111/j.1553-2712.2008.00188.x](https://doi.org/10.1111/j.1553-2712.2008.00188.x).
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65(9 Suppl):S63-7.
- Epstein RM. Assessment in medical education. *N Engl J Med* 2007;356:387-96, doi:[10.1056/NEJMe078002](https://doi.org/10.1056/NEJMe078002).
- Sherbino J, Bandiera G, Frank J. Assessing competence in emergency medicine trainees; an overview of effective methodologies. *Can J Emerg Med* 2008;10:365.
- Lammers RL, Davenport M, Korley F, et al. Teaching and assessing procedural skills using simulation: metrics and methodology. *Acad Emerg Med* 2008;15:1079-87, doi:[10.1111/j.1553-2712.2008.00233.x](https://doi.org/10.1111/j.1553-2712.2008.00233.x).
- Boulet JR. Summative assessment in medicine: the promise of simulation for high-stakes evaluation. *Acad Emerg Med* 2008;15:1017-24, doi:[10.1111/j.1553-2712.2008.00228.x](https://doi.org/10.1111/j.1553-2712.2008.00228.x).
- Spillane L, Hayden E, Fernandez R, et al. The assessment of individual cognitive expertise and clinical competency: a research agenda. *Acad Emerg Med* 2008;15:1071-8, doi:[10.1111/j.1553-2712.2008.00271.x](https://doi.org/10.1111/j.1553-2712.2008.00271.x).
- Okuda Y, Bryson EO, DeMaria S Jr, et al. The utility of simulation in medical education: what is the evidence? *Mt Sinai J Med* 2009;76:330-43, doi:[10.1002/msj.20127](https://doi.org/10.1002/msj.20127).
- Fritz PZ, Gray T, Flanagan B. Review of mannequin-based high-fidelity simulation in emergency medicine. *Emerg Med Australas* 2008;20:1-9, doi:[10.1111/j.1742-6723.2007.01022.x](https://doi.org/10.1111/j.1742-6723.2007.01022.x).
- Okuda Y, Bond W, Bonfante G, et al. National growth in simulation training within emergency medicine residency programs, 2003-2008. *Acad Emerg Med* 2008;15:1113-6, doi:[10.1111/j.1553-2712.2008.00195.x](https://doi.org/10.1111/j.1553-2712.2008.00195.x).
- Binstadt ES, Walls RM, White BA, et al. A comprehensive medical simulation education curriculum for emergency medicine residents. *Ann Emerg Med* 2007;49:495, doi:[10.1016/j.annemergmed.2006.08.023](https://doi.org/10.1016/j.annemergmed.2006.08.023).
- Philibert I. Accreditation Council for Graduate Medical Education. *ACGME Bulletin*. December 2005. Available at: www.acgme.org (accessed April 2010).
- Adler MD, Trainor JL, Siddall VJ, et al. Development and evaluation of high-fidelity simulation case scenarios for pediatric resident education. *Ambul Pediatr* 2007;7:182-6, doi:[10.1016/j.ambp.2006.12.005](https://doi.org/10.1016/j.ambp.2006.12.005).
- Brett-Fleegler MB, Vinci RJ, Weiner DL, et al. A simulator-based tool that assesses pediatric resident resuscitation competency. *Pediatrics* 2008;121:e597-603, doi:[10.1542/peds.2005-1259](https://doi.org/10.1542/peds.2005-1259).
- Morgan PJ, Cleave-Hogg DM, Guest CB, et al. Validity and reliability of undergraduate performance assessments in an anesthesia simulator. *Can J Anaesth* 2001;48:225-33, doi:[10.1007/BF03019750](https://doi.org/10.1007/BF03019750).
- Murray DJ, Boulet JR, Avidan M, et al. Performance of residents and anesthesiologists in a simulation-based skill assessment. *Anesthesiology* 2007;107:705-13, doi:[10.1097/01.anes.0000286926.01083.9d](https://doi.org/10.1097/01.anes.0000286926.01083.9d).
- Weller JM, Bloch M, Young S, et al. Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. *Br J Anaesth* 2003;90:43-7, doi:[10.1093/bja/90.1.43](https://doi.org/10.1093/bja/90.1.43).
- Berkenstadt H, Ziv A, Gafni N, et al. Incorporating simulation-based objective structured clinical examination into the Israeli National Board Examination in Anesthesiology. *Anesth Analg* 2006;102:853-8, doi:[10.1213/01.ane.0000194934.34552.ab](https://doi.org/10.1213/01.ane.0000194934.34552.ab).
- Girzadas DV Jr, Clay L, Caris J, et al. High fidelity simulation can discriminate between novice and experienced residents when assessing competency in patient care. *Med Teach* 2007;29:452-6, doi:[10.1080/01421590701513698](https://doi.org/10.1080/01421590701513698).
- Dagnone JD. Assessment tool for a standardized ventricular fibrillation cardiac arrest. *Can J Emerg Med* 2009;11:262.
- Andreatta PB, Gruppen LD. Conceptualising and classifying validity evidence for simulation. *Med Educ* 2009;43:1028-35, doi:[10.1111/j.1365-2923.2009.03454.x](https://doi.org/10.1111/j.1365-2923.2009.03454.x).
- Walls RM. Airway. In: Marks JA, editor. *Rosen's emergency medicine*. Philadelphia: Mosby Elsevier; 2006. p. 2-26.
- Morgan PJ, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Acad Med* 2001;76:1053-5, doi:[10.1097/00001888-200110000-00016](https://doi.org/10.1097/00001888-200110000-00016).
- Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73:993-7, doi:[10.1097/00001888-199809000-00020](https://doi.org/10.1097/00001888-199809000-00020).