

# The McGill Simulation Complexity Score (MSCS): a novel complexity scoring system for simulations in trauma

Kosar Khwaja, MD, MBA, MSc  
 Melina Deban, MD CM  
 Sameena Iqbal, MD  
 Jalal Alowais, MD  
 Bader Al Bader, MD  
 Dan Deckelbaum, MD  
 Tarek Razek, MD CM

Accepted Aug. 12, 2020

## Correspondence to:

Kosar Khwaja  
 Departments of Surgery and Critical Care  
 Medicine  
 McGill University Health Centre  
 1650 Cedar Ave., L9.411  
 Montréal QC H3G 1A4  
 dr.k.khwaja@mcgill.ca

**Cite as:** *Can J Surg* 2023 April 21; 66(2).  
 doi: 10.1503/cjs.002220

**Background:** In medical education, simulation can be defined as an activity in which an individual demonstrates skills, procedures and critical thinking using interactive mannequins in a setting closely resembling the clinical environment. To our knowledge, the complexity of trauma simulations has not previously been assessed. We aimed to develop an objective trauma simulation complexity score and assess its interrater reliability.

**Methods:** The McGill Simulation Complexity Score (MSCS) was designed to address the need for objective evaluation of the complexity of trauma scenarios. Components of the score reflected the Advanced Trauma Life Support approach to trauma. The score was developed to take into account the severity of trauma injuries and the complexity of their management. We assessed interrater reliability at 5 high-fidelity simulation events. Interrater reliability was calculated using the Pearson correlation coefficient (PCC) and the intraclass correlation coefficient (ICC).

**Results:** The MSCS has 5 categories: airway, breathing, circulation, disability, and extremities or exposure. The scale has 5 levels for each category, from 0 to 4; level increases with complexity, with 0 corresponding to normal or absent. Cases designed to lead to cardiac arrest, regardless of whether or not the trainee has the ability to resuscitate the simulated patient and regardless of the level of each category, are automatically assigned the maximum score. Between 3 and 9 raters used the MSCS to grade the level of complexity of 26 scenarios at the 5 events. The mean MSCS was 10.2 (range 3.0–20.0). Mean PCC and ICC values were both above 0.7 and therefore statistically significant.

**Conclusion:** The MSCS for trauma is an innovative scoring system with high interrater reliability.

**Contexte :** Dans le contexte de la formation médicale, la simulation peut se définir comme une activité au cours de laquelle une personne démontre ses habiletés, pose des gestes et exerce son jugement critique en se servant de mannequins interactifs dans un environnement qui imite de près la réalité clinique. À notre connaissance, la complexité des simulations en traumatologie n'avait encore jamais été évaluée. Nous avons voulu mettre au point un indice de complexité objectif des simulations en traumatologie et en évaluer la fiabilité interévaluateurs.

**Méthodes :** L'indice McGill de complexité des simulations (IMCS) a été conçu pour répondre au besoin d'évaluer objectivement la complexité des scénarios en traumatologie. Les éléments de l'indice faisaient écho à l'approche dite des soins avancés de réanimation traumatologique. L'indice a été conçu de manière à tenir compte de la gravité des traumatismes et de la complexité de leur prise en charge. Nous avons mesuré la fiabilité interévaluateurs de 5 événements simulés en haute-fidélité. La fiabilité interévaluateurs a été calculée à l'aide du coefficient de corrélation de Pearson (CCP) et du coefficient de corrélation intraclass (CCI).

**Résultats :** L'IMCS compte 5 catégories : voies respiratoires, respiration, circulation, incapacité et membres ou exposition. Pour chaque catégorie, il y a une échelle à 5 niveaux allant de 0 à 4; le niveau augmente en fonction de la complexité, 0 correspondant à une situation normale ou à l'absence de problème. Les scénarios conçus pour mener à l'arrêt cardiaque, indépendamment de la capacité du stagiaire à procéder à une réanimation sur le mannequin et indépendamment de chaque catégorie, reçoivent d'emblée le score maximum. De 3 à 9 évaluateurs ont utilisé l'IMCS pour classer le niveau de complexité de 26 scénarios pour les 5 événements. L'IMCS moyen a été de 10,2 (écart 3,0–20,0). Le CCP et le CCI ont tous deux été supérieurs à 0,7 et donc statistiquement significatifs.

**Conclusion :** L'IMCS en traumatologie est un système de classification novateur doté d'une grande fiabilité interévaluateurs.

In medical education, simulations are increasingly being used for clinical skill practice and evaluation. Simulation can be defined as an activity in which an individual demonstrates skills, procedures and critical thinking using interactive mannequins in a setting closely resembling the clinical environment.<sup>1</sup> As an evaluation tool, it integrates all 4 components of Miller's model of clinical assessment, namely knowledge (knows), competence (knows how), performance (shows how) and action (does).<sup>2</sup>

Different types of simulations, including high-fidelity simulation, are used in trauma surgery.<sup>3</sup> High-fidelity simulations involve situations that closely resemble reality, in which the physical inputs are highly realistic and there is a high degree of interactivity with the trainee.<sup>1</sup>

To our knowledge, complexity assessment has never been explored in the context of trauma simulation and there exists no score objectively evaluating the level of complexity of a trauma simulation scenario; existing scores were all developed for real patients. The American Society of Anesthesiologists (ASA) physical status classification system groups patients according to their illness severity with the goal of assessing their anesthetic risk.<sup>4</sup> The Charlson Comorbidity Index (CCI) predicts the 10-year mortality of patients on the basis of their comorbidities.<sup>5</sup> The Injury Severity Score (ISS) describes the severity of injury of patients who have sustained multiple trauma and correlates with survival.<sup>6</sup> Although the ISS is applicable to trauma, it does not allow stratification of the difficulty of managing trauma patients in the context of a high-fidelity simulation. To our knowledge, a method does not currently exist to develop scenarios of standard difficulty that match trainees' skill levels, which can be used for teaching and evaluation. Our objective was to design and assess the reliability of a score that could be used in such a process. This article describes and validates the McGill Simulation Complexity Score (MSCS).

## METHODS

The MSCS was designed to address the need for objective evaluation of complexity in trauma scenarios. Components of the score reflect the Advanced Trauma Life Support (ATLS) approach to trauma. The score was designed to reflect the severity of trauma injuries and the complexity of their management.

The MSCS was presented to surgeons with expertise in trauma and simulation at a national simulation course, for feedback on its applicability and face validity.

We aimed to assess the interrater reliability of the MSCS at 5 high-fidelity simulation events. Of these 5 events, 2 were in the context of military training, 2 were in the context of postgraduate training and the last was in the context of a national simulation course. The events took place in Montréal and Calgary, Canada.

The cases were designed by the event organizers, independently from the raters. Senior general surgery residents, trauma fellows and staff present at the event were offered the opportunity to rate the simulation scenarios using the MSCS grid. All participation was voluntary. Before the simulation, the scenarios were reviewed with all of the raters to ensure they had a thorough and uniform knowledge of the case. They subsequently proceeded to score the scenarios independently. No form of communication was allowed during the scoring step. Mannequin-based simulations took place after the reviewing and scoring processes. McGill University's institutional review board approved our study protocol.

## Statistical analysis

Interrater reliability was assessed using the Pearson correlation coefficient (PCC) and intraclass correlation coefficient (ICC). Both were calculated with SPSS software (version 20.0.0, IBM Corp.). Significance was deemed to be reached at a cut-off value of 0.7.<sup>7</sup> A standard deviation (SD) was included for descriptive purposes. Analysis for each event was carried out separately, given that cases and raters were different across events.

## RESULTS

The MSCS components are based on the ATLS approach to trauma, with 5 categories: airway, breathing, circulation, disability, and extremities or exposure (Figure 1). The scale has 5 levels for each category, from 0 to 4; level increases with complexity, with 0 corresponding to normal or absent. Application of the score is clarified through bolding of the main criteria. Nonexhaustive examples that match each level of complexity are provided. MSCS values can range from 0 (easy) to 20 (very difficult). Cases designed to lead to cardiac arrest, regardless of whether or not the trainee has the ability to resuscitate the patient and regardless of the level of each category, automatically get the maximum score.

None of the people invited to participate in the rating exercise declined. Raters used the MSCS to grade the level of complexity of 26 scenarios that took place across 5 separate events; each scenario was assessed by 3–9 raters. The mean MSCS was 10.2 (range 3.0–20.0) (Table 1). The mean PCC and ICC values were both above 0.7 and therefore considered to be statistically significant.<sup>7</sup> The SD was 0.334 (range 0.001–0.661), meaning that the ratings differed, on average, by 0.33 points for each component of the score. Floor and ceiling effects were observed, as there was no variation among raters for the 2 easiest scenarios (MSCS 3) and the 4 most difficult scenarios (MSCS 18.4 and 20) (Figure 2).

Evaluator (name, level): _____		Simulation Score			
Scenario: _____					
Simulation Score	Airway	Breathing	Circulation	Deficits	Extremities/Exposure
<b>0</b>	<b>A0:</b> Normal and patent Airway	<b>B0:</b> Normal breathing	<b>C0:</b> Normal Circulation	<b>D0:</b> No neurologic deficits	<b>E0:</b> Normal extremities
<b>1</b>	<b>A1:</b> Airway <b>patent</b> • <b>jaw thrust</b> • <b>Bag mask</b> ventilation or • <b>easy intubation</b> (grade 1) ○ altering mental status • intubation prior to transfer	<b>B1:</b> <b>Mild</b> deficits in breathing • <b>Tachypnea</b> (RR ≥ 25) • accessory muscle use • <b>no evidence of hypoxia</b> (compensated)	<b>C1:</b> <b>Mild</b> deficit in circulation • HR > 110 • SPB > 100 • <b>response to &lt; 2L</b> crystalloid infusion	<b>D1:</b> <b>Mild</b> head injuries • <b>GCS 14–15</b> <b>Brachial plexus</b> injury • neuro deficits	<b>E1:</b> <b>Closed</b> extremity fractures • <b>without vascular compromise</b> <b>1<sup>st</sup> degree</b> burn
<b>2</b>	<b>A2:</b> <b>Moderate</b> difficult airway • <b>laryngoscopy</b> possible (grade 2–3), e.g. ○ blood in oral-pharynx ○ subcutaneous emphysema ○ deviated trachea ○ neck hematoma	<b>B2:</b> <b>Moderate</b> deficits • hypoxia • <b>resolution with simple procedures</b> , i.e. chest tube • <b>simple pneumothorax</b> • <b>flail chest</b> • subcutaneous emphysema	<b>C2:</b> <b>Moderate</b> evidence of circulatory shock • hypotension <b>responding to blood product</b> transfusion (<10 u RBC) • major external bleeding <b>amenable to direct pressure</b>	<b>D2:</b> <b>Moderate</b> head injuries • <b>GCS 9–13</b> <b>Thoracic or lumbar</b> spine injury • neuro deficits	<b>E2:</b> <b>Open</b> extremity fractures or any extremity injury • <b>with vascular compromise</b> <b>2<sup>nd</sup> degree</b> burn • < 9% of BSA
<b>3</b>	<b>A3:</b> <b>Difficult</b> airway • <b>expert oral-endotracheal intubation</b> (with fiberoptic or glidescope) • <b>surgical airway</b> , e.g. ○ major facial fractures ○ blood in airway ○ enlarged tongue ○ obese ○ short neck	<b>B3:</b> • <b>Hemodynamically unstable</b> (obstructive shock) • <b>tension pneumothorax</b> (with or without hypoxia) • significant subcutaneous emphysema • <b>fluctuations</b> in hypoxia	<b>C3:</b> <b>Significant</b> circulatory shock • <b>transient</b> response • massive transfusion • necessitating • <b>angio/operative</b> source control	<b>D3:</b> <b>Severe</b> head injury • GCS < 8 <b>High cervical</b> spine injuries • neuro deficits • spinal shock • with or without respiratory failure	<b>E3:</b> Major extremity injury • with <b>amputation</b> <b>2<sup>nd</sup> degree</b> burn • <b>9%-18%</b> of body surface area <b>3<sup>rd</sup> degree</b> burn • < 9% of body surface area
<b>4</b>	<b>A4:</b> Major airway obstruction • necessitating <b>immediate surgical airway</b> , e.g. ○ unable to intubate orally ○ penetrating airway injury ○ blunt laryngeal fracture	<b>B4:</b> • <b>Persistent</b> significant hypoxia • major pulmonary <b>contusions</b> • <b>air embolism</b> • <b>TRALI</b>	<b>C4:</b> <b>Profound</b> sustained circulatory shock • despite appropriate resuscitation and maneuvers ( <b>non-responder</b> ) • necessitating <b>operative</b> source control	<b>D4:</b> <b>Severe</b> head injury • <b>pupil dilation</b> <b>Severe</b> spine injury • neuro deficits • spinal shock • <b>hypotension and bradycardia</b>	<b>E4:</b> <b>Mangled</b> extremity • with or without vascular compromise • significant <b>degloving</b> injury <b>2<sup>nd</sup> degree</b> burn • > 18% of BSA <b>3<sup>rd</sup> degree</b> burn • > 9% of BSA
<b>Score</b>					
<b>Total Score</b>					

NB: bolded criteria should mainly determine level of complexity.  
**Any scenario that is designed to proceed to a cardiac arrest with or without the ability to resuscitate or death should receive a score of 20.**

**Fig. 1.** Determination of the McGill Simulation Complexity Score. Note: BSA = body surface area, GCS = Glasgow Coma Scale, HR = heart rate, RBC = red blood cells, RR = respiratory rate, SPB = systolic blood pressure, TRALI = transfusion-related acute lung injury.

**DISCUSSION**

We created the McGill Simulation Complexity Score because of our need to objectively evaluate the complexity of simulation scenarios. Furthermore, we felt that there was a lack of an appropriate tool to compare simulation scenarios and ensure that trainees received uniform exposure to scenarios with different levels of difficulty across simulation events.

The MSCS takes into account all 5 components of the ATLS approach to trauma. Cases are scored on the basis of the initial presentation of the simulated patient. In other words, the score should be based on the initial situation, regardless of how the case evolves. The only exception is a scenario designed to lead to a cardiac arrest, in which case the scenario is automatically scored as a 20 (maximum score). There was considerable discussion

around this aspect of the score. Some raters felt that a case leading to a cardiac arrest should be assigned a lower score, given the standard approach to such an event. However, after substantial deliberation, it was decided that a trauma scenario designed a priori to lead to a cardiac arrest and its ensuing management, regardless of whether the scenario involves a blunt or penetrating trauma, can be extremely challenging and overwhelming for trainees and should automatically be scored as a 20. Such scenarios often create substantial anxiety and require quick action. They are typically designed for senior-level trainees and experts in traumatology.

During the face validity process, raters noted that certain elements of a scenario could be classified under more than 1 category within the ATLS construct. An example of such a situation is a tension pneumothorax, which has a breathing and circulation (shock) component. To avoid a

double evaluation of tension pneumothorax within the breathing and the circulation categories, tension pneumothorax was listed in section B3 of the scoring sheet (a breathing problem leading to hemodynamic instability), as shown in Figure 1. We refrained from including too many details on the scoring sheet (e.g., description of burns in airway) because we wanted to keep the MSCS simple and easy to use.

High interrater reliability was demonstrated through statistically significant PCC and ICC values. We calculated PCC to compare the relative values of each component of the score, and we calculated ICC to correlate the absolute values of the components of the score. In other words, the high PCC indicates that the raters had a uniform understanding of the scenarios. They were able to identify which component of the scenario was challenging. For example, if circulation was a difficult aspect of the case, a high PCC shows that relative to airway,

breathing, disability and exposure, raters gave a higher score to circulation. Ratings of A1, B1, C3, D1 and E0 and of A2, B2, C4, D2 and E1 by 2 different raters would yield an elevated PCC. Although the overall MSCS value differed for the 2 raters, the relative difference between the components of the MSCS is similar (i.e., raters agreed that the C component was more challenging than the A, B, D and E components).

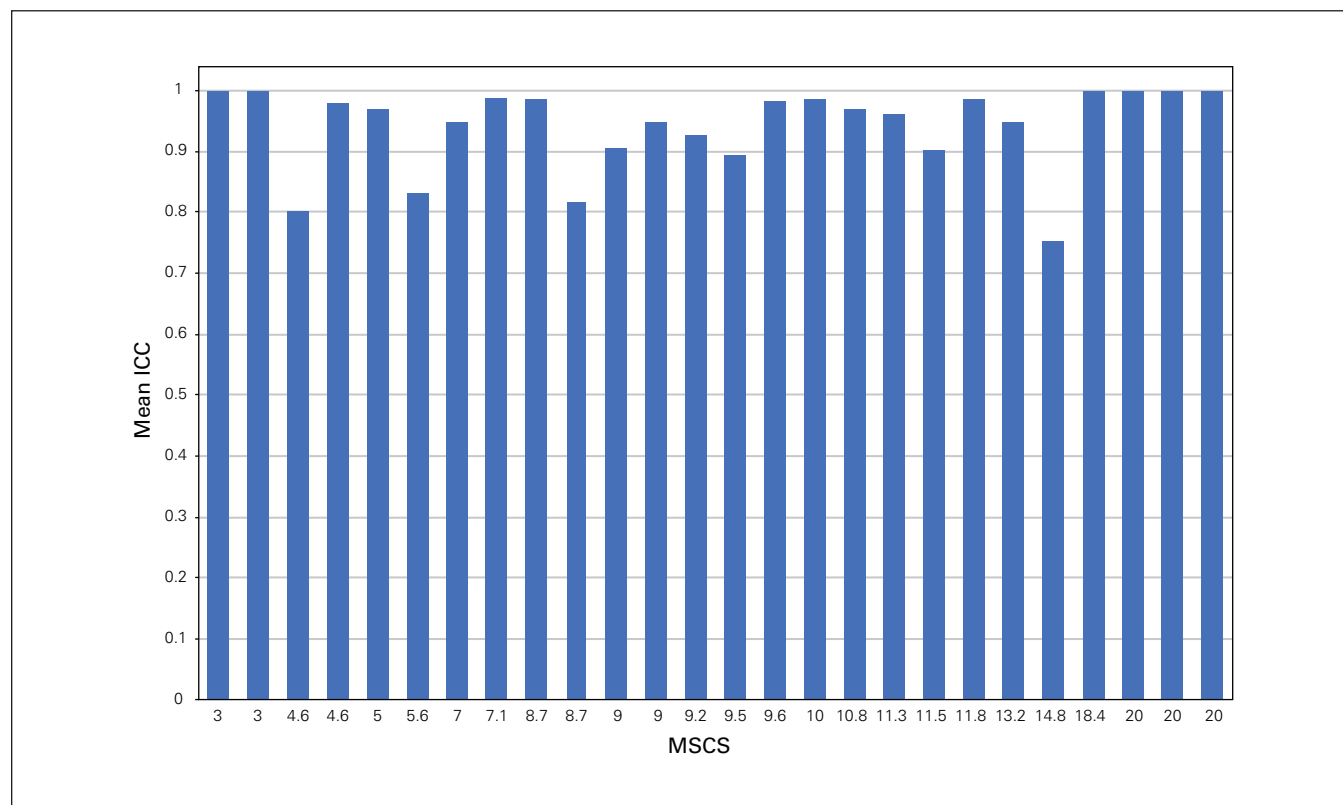
In this context, the ICC is useful for comparing the absolute values of components. A high ICC can only be the result of ratings such as A1, B2, C1, D0 and E4 and of A1, B2, C1, D0 and E4; that is, the absolute values of the scores for components A, B, C, D and E of rater 1 equal the absolute values of the scores for components A, B, C, D and E of rater 2. Therefore, an elevated ICC shows that raters agreed on how to rate a scenario using the MSCS.<sup>8</sup>

The SD was calculated for descriptive purposes. It shows how much variation there is among scores for each component of the MSCS. The average value was small (0.334), further supporting the high interrater reliability we found. In other words, the MSCS was clear enough, or reliable enough, to yield similar ratings when used by different raters blinded to each other's scoring.

Furthermore, floor and ceiling effects were observed for the MSCS ratings for all 26 scenarios. Indeed, there was minimal or no variation in the ICC at the extremes

**Table 1. Total McGill Simulation Complexity Score and interrater reliability statistics for the rating exercise**

Measure	Mean (range)
Total McGill Simulation Complexity Score	10.2 (3.0–20.0)
Pearson correlation coefficient	0.979 (0.879–1.000)
Intraclass correlation coefficient	0.942 (0.752–1.000)
Standard deviation	0.334 (0.000–0.661)



**Fig. 2.** Mean intraclass correlation coefficient (ICC) of 26 mannequin-based simulation scenarios in order of increasing McGill Simulation Complexity Score (MSCS) values.

of the score. This effect indicates how straightforward it is to rate scenarios at the extremes of the MSCS; it is easy to recognize the easy and very difficult scenarios. We must recognize that an automatic score of 20 in the context of cardiorespiratory arrest can contribute to the ceiling effect. However, for scores near 0 and 19, this effect can be explained by the simple nature of scenarios situated at the extremes. Floor and ceiling effects can be negative attributes of a score, as they can indicate the inability of the score to distinguish features of cases located at the extremes.<sup>9</sup> In our case, this effect was restricted to very low (3) and very high (18–20) scores; consequently, we do not estimate that there is extensive blunting of scoring at the poles of complexity. On the contrary, the presence of both effects leads us to conclude that the score captures the whole range of complexity of trauma scenarios.

Simulation training has gained popularity because it offers numerous advantages. On one hand, the realism of high-fidelity scenarios provides an environment that promotes coordination and communication under stress.<sup>10</sup> On the other hand, the fictional nature of the exercise guarantees that no patients will be harmed. Simulation training therefore provides a relatively stress-free learning experience for trainees, making it a cost-effective way to reduce human errors when it comes to treating patients in hospital.<sup>11</sup> Beyond its advantages with respect to learning, simulation also facilitates evaluation of a unique component of trainee performance: competency. Del Bueno and colleagues introduced the concept of competency as a multifaceted set of aptitudes that goes beyond simple knowledge acquisition and includes technical skills, critical thinking and interpersonal skills, all of which are demonstrated in high-fidelity exercises.<sup>12</sup>

Gordon and colleagues reported that more than three-quarters of their study participants (residents, fellows and medical students) perceived their performance on a simulator to be more representative of their capacities than their performance on an oral objective structured clinical examination.<sup>13</sup> Although high-fidelity simulation is far from the point where it would be able to replace oral examinations for many certification boards (American Board of Surgery, Royal College of Physicians and Surgeons of Canada), the MSCS now enables us to objectively quantify the level of difficulty of scenarios that may be used to evaluate trainee performance. The MSCS could help educators to determine the level of complexity best suited for various types and levels of trainees (nurses, medical students, residents, fellows, physician assistants) and then standardize the complexity level of scenarios presented to trainees at different sites. The MSCS can therefore facilitate the design of scenarios with a predetermined level of complexity for learning and performance evalu-

ation purposes. If the appropriate level of complexity for a given learner can be identified, educators can optimize the simulated case by avoiding oversimplifying or overcomplicating the scenario.

The MSCS can also be used to address gaps in knowledge. If cases are built using the MSCS framework, the level at which the trainee displays difficulty is easily identifiable. For instance, if a trainee struggles in a scenario of level A3, B1, C1, D0 and E0, then teaching objectives should include an emphasis on expert orotracheal intubation and surgical airway. With a more structured case design, knowledge deficits are more easily identified.

Complexity assessment could become part of the overall assessment of scenario quality. As mentioned previously, the MSCS can highlight knowledge deficits, but it can also clarify learning objectives. The score can enable scenario creators to ensure that the design of a given scenario is complete, by checking that each trauma component, namely airway, breathing, circulation, disability and exposure, has been addressed and is adequately assessed.

The MSCS validation process included 2 military simulation sessions conducted in Montréal, Canada. We see a substantial opportunity to use the MSCS in the creation of standardized military trauma simulation scenarios that can then be deployed at various military training sites. Although we did not use the MSCS to assess mass casualty scenarios in this study, scenario creators could use the MSCS framework to quantify the numbers of patients with various types of complex injuries, to ensure that their scenario accurately represents the array of patients with complex injuries that care providers might face during a mass casualty event.

### *Limitations*

Not having the same raters across the events was a limitation of this study. It was challenging, in terms of logistics, to coordinate a single set of raters to be present at all of the simulation events. If we had been able to do this, we would have been able to ensure that all raters had similar learning curves, a parameter that we did not measure in this study. This limitation was, however, offset by the number of scenarios, namely 26, that were rated in total. Even though there were logistical barriers to recruiting more raters or keeping the same ones for each event, this situation reflects real life, and our results highlight the fact that the score can be applied in different settings by different raters with high reliability. Another limitation was that the score was applied retrospectively to scenarios that had already been built. Further studies will be conducted with simulation scenarios built prospectively on the basis of the MSCS framework.

## CONCLUSION

The MSCS for trauma is an innovative scoring system that quantifies the complexity of trauma simulation scenarios. High interrater reliability was observed among raters using the MSCS to rate multiple high-fidelity simulation scenarios. The MSCS is an easy-to-use tool that can allow scenarios used for training and trainee performance evaluation to be compared in terms of their complexity.

**Affiliations:** From the Departments of Surgery (Khawaja, Deban, Alowais, Al Bader, Deckelbaum, Razek) and Critical Care Medicine (Khawaja), McGill University Health Centre, Montréal, Que. ; and the Department of Nephrology, McGill University Health Centre, Montréal, Que. (Iqbal).

**Competing interests:** None declared.

**Contributors:** K. Khawaja, M. Deban, D. Deckelbaum and T. Razek designed the study. M. Deban, Jalal Alowais and B. Al Bader acquired the data, which M. Deban, S. Iqbal and B. Al Bader analyzed. M. Deban and S. Iqbal, wrote the article, which all authors critically revised. All authors gave approval of the final version to be published.

**Content licence:** This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

1. Jeffries PR. A framework for designing, implementing, and evaluating simulations used as teaching strategies in nursing. *Nurs Educ Perspect* 2005;26:96-103.
2. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65:S63-7.
3. Lee SK, Pardo M, Gaba D, et al. Trauma assessment training with a patient simulator: a prospective, randomized study. *J Trauma* 2003;55:651-7.
4. Daabiss M. American Society of Anaesthesiologists physical status classification. *Indian J Anaesth* 2011;55:111-5.
5. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373-83.
6. Baker SP, O'Neil B, Haddon W, et al. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma* 1974;14:187-96.
7. Guyatt G, Rennie D, Meade MO, et al. *Users' guide to the medical literature*. 2nd ed. Chicago: JAMA Press, 2002.
8. Osborne JW. *Best practices in quantitative methods*. London: Sage Publications Inc., 2007.
9. Cramer D, Howitt DL. *The SAGE dictionary of statistics: a practical resource for students in the social sciences*. 3rd ed. London: Sage Publications Inc., 2004.
10. Small SD, Wuerz RC, Simon R, et al. Demonstration of high-fidelity simulation team training for emergency medicine. *Acad Emerg Med* 1999;6:312-23.
11. McConnell H, Pardy A. Virtual patient simulation for prevention of medical error: beyond just technical upskilling. *World Hosp Health Serv* 2008;44:36-9.
12. Del Bueno DJ, Barker F, Christmyer C. Implementing a competency-based orientation program. *Nurse Educ* 1980;5:16-20.
13. Gordon JA, Tancredi DN, Binder WD, et al. Assessment of a clinical performance evaluation tool for use in a simulator-based testing environment: a pilot study. *Acad Med* 2003;78:S45-7.