

Harnessing the power of simulation for assessment: Consensus recommendations for the use of simulation-based assessment in emergency medicine

Andrew K. Hall , MD, MMed*[†]; Timothy Chaplin, MD*[‡]; Tamara McColl, MD[‡]; Andrew Petrosoniak, MD, MSc (Med Ed)^{§¶}; Kyla Caners, MD^{||}; Nicole Rocca, MD, MSc**[‡]; Carlyn Gardner, MD^{††}; Farhan Bhanji, MD, MSc^{††§§}; Rob Woods , MD, MMed^{††}

ABSTRACT

Objectives: To address the increasing demand for the use of simulation for assessment, our objective was to review the literature pertaining to simulation-based assessment and develop a set of consensus-based expert-informed recommendations on the use of simulation-based assessment as presented at the 2019 Canadian Association of Emergency Physicians (CAEP) Academic Symposium on Education.

Methods: A panel of Emergency Medicine (EM) physicians from across Canada, with leadership roles in simulation and/or assessment, was formed to develop the recommendations. An initial scoping literature review was conducted to extract principles of simulation-based assessment. These principles were refined via thematic analysis, and then used to derive a set of recommendations for the use of simulation-based assessment, organized by the Consensus Framework for Good Assessment. This was reviewed and revised via a national stakeholder survey, and then the recommendations were presented and revised at the consensus conference to generate a final set of recommendations on the use of simulation-based assessment in EM.

Conclusion: We developed a set of recommendations for simulation-based assessment, using consensus-based expert-informed methods, across the domains of validity, reproducibility, feasibility, educational and catalytic effects, acceptability, and programmatic assessment. While the precise role of simulation-based assessment will be a subject of continued debate, we propose that these recommendations be used to assist educators and program leaders as they incorporate simulation-based assessment into their programs of assessment.

RÉSUMÉ

Objectifs: Compte tenu du recours de plus en plus fréquent à la simulation aux fins d'évaluation, l'étude visait à procéder à un examen de la documentation sur l'évaluation fondée sur la simulation et à élaborer un ensemble de recommandations obtenues par consensus et reposant sur l'avis d'experts sur l'application de l'évaluation fondée sur la simulation, ensemble qui a été présenté à l'occasion du Symposium 2019 de la section des affaires universitaires de l'Association canadienne des médecins d'urgence, sur la formation.

Méthode: Un groupe de médecins d'urgence (MU) provenant de partout au Canada et reconnus pour jouer des rôles de premier plan en simulation ou en évaluation a été formé afin d'élaborer des recommandations. Il y a d'abord eu un examen de la documentation afin d'en délimiter l'étendue et de dégager les grands principes de l'évaluation fondée sur la simulation. Ceux-ci ont par la suite été affinés à l'aide d'une analyse thématique, puis ont servi à l'élaboration d'un ensemble de recommandations aux fins de l'évaluation fondée sur la simulation, structuré selon le Consensus Framework for Good Assessment. La version provisoire a ensuite été soumise, à l'aide d'une enquête menée à l'échelle nationale, à l'examen des parties intéressées et a fait l'objet de modifications. Enfin, les recommandations ont été présentées, puis modifiées à l'occasion du congrès de consensus, en vue de l'élaboration d'un ensemble définitif de recommandations sur l'application de l'évaluation fondée sur la simulation, en MU.

Conclusion: Le processus a permis d'élaborer un ensemble de recommandations sur l'application de l'évaluation fondée sur la simulation, obtenues par consensus et reposant sur l'avis

From *Department of Emergency Medicine, Queen's University; [†]Royal College of Physicians and Surgeons of Canada; [‡]Department of Emergency Medicine, University of Manitoba; [§]Department of Medicine, University of Toronto; [¶]Li Ka Shing Knowledge Institute, St. Michael's Hospital; ^{||}Department of Medicine, Division of Emergency Medicine, McMaster University; ^{**}Departments of Emergency Medicine & Critical Care Medicine, Queen's University; ^{††}Department of Emergency Medicine, University of Saskatchewan; ^{‡‡}Department of Pediatrics, McGill University; ^{†††}Royal College of Physicians and Surgeons of Canada; ^{§§}Steinberg Centre for Simulation and Interactive Learning; and the ^{††††}Department of Emergency Medicine, University of Saskatchewan.

Correspondence to: Dr. Timothy Chaplin, Queen's University, Department of Emergency Medicine, Kingston Health Sciences Centre, 76 Stuart Street, Kingston, Ontario K7L 2V7; Email: t.chaplin@queensu.ca

d'experts, dans les domaines de la validité, de la reproductibilité, de la faisabilité, des effets éducatifs et « catalytiques », de l'acceptabilité et de l'évaluation programmatique. Bien que le rôle précis de l'évaluation fondée sur la simulation fasse encore l'objet de débats, le groupe propose d'appliquer ces recommandations afin d'aider les éducateurs et les

responsables de programmes à mesure que l'évaluation fondée sur la simulation s'intègre dans les modalités d'évaluation.

Keywords: Education, education research, emergency medicine, simulation

INTRODUCTION

Facing concerns about deficiencies in current educational and assessment processes, medical education is evolving rapidly to meet the complex needs of our learners, administrative organizations, and the public.¹ Specifically, the transition to competency-based medical education has placed an increased onus on training programs and systems to ensure that graduating trainees meet key competencies before entering independent practice.² To achieve this, education leaders have proposed an increase in the use of direct workplace-based observation to assure that these competencies are met.³ However, in emergency medicine (EM) and many other clinical workplaces, experiences are not predictable. This results in significant variability across trainees in the competencies that may be opportunistically assessed.⁴ Further, the high acuity inherent to EM requires that patient care be prioritized over trainee assessment.

Simulation-based assessment has been proposed as a potential solution to this problem,⁵ with the capacity to control exposure to scheduled reproducible experiences and allow trainees to demonstrate their abilities without any risk to patient safety.⁶ From occasional low-stakes assessment to higher-stakes examination, there has been increasing use of simulation-based assessment across medical specialties.⁷ In Canadian EM, few training programs had implemented simulation-based assessment by 2017,⁸ but this number is steadily rising. There is literature supporting the translational outcomes (e.g., improved performance and patient care) of simulation-based training^{9,10} but limited evidence directly correlating simulation-based assessment to real-world performance.¹¹ While, in principle, the simulated environment seems ideal for assessment, there are several tensions or concerns that stakeholders have raised. For example, simulation was initially developed as a “safe space” for practice,¹² and the introduction of assessment may threaten the

integrity of this learning environment, with trainees fearing negative assessment. Another concern is variable access to simulation equipment and how this may disadvantage trainees and programs with resource limitations.

Despite these tensions, program directors and educators have been tasked with the rapid integration of simulation-based assessment into programs of assessment without a clear understanding of how best to use it effectively. To assist those responsible for the implementation of simulation-based assessment, it was an aim of the 2019 Canadian Association of Emergency Physicians (CAEP) Academic Symposium on Education to review the literature pertaining to simulation-based assessment and develop a set of consensus-based and expert-informed recommendations on the use of simulation for assessment. In this paper, we describe these recommendations for simulation-based assessment, organized by the Consensus Framework for Good Assessment of Norcini et al.,¹³ to assist educators in the era of competency-based medical education.

METHODS

In May 2019, the Academic Section of the CAEP held its annual consensus conference on education. In preparation for this conference, a working group of emergency physicians and content experts was formed with representation from multiple institutions across Canada. Members were chosen based on their experience and expertise in medical simulation, assessment, or both, having completed degrees in medical education or advanced training in simulation. Monthly teleconference meetings were held to design and implement the study. To obtain a final product of consensus-based recommendations for simulation-based assessment, a scoping literature review was first performed to identify principles of simulation-based assessment. These principles were used to derive a set of recommendations that were revised using a

national stakeholder survey. Finally, they were presented and revised at the consensus conference to generate a final set of recommendations on the use of simulation-based assessment in EM. This study received approval from the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (REB #6023280)

Scoping review

With the aid of a university librarian, we performed a scoping review¹⁴ to collate existing literature on the use of simulation-based assessment in medical education. MEDLINE, EMBASE, and ERIC databases were searched, along with Google Scholar, in October 2018 using “and/or” combinations of the following keywords: “simulation,” “manikins,” “assessment,” “competence,” “residency,” and “medical education.” All searches were limited to English-language papers published in or after the year 2000. All papers were initially screened for inclusion by title and, then, abstract by two study investigators (AKH and TC). Papers were included if they either: 1) made recommendations on the use of technology-enhanced simulation for assessment in medical education; or 2) investigated the use of technology-enhanced simulation to assess physician learners at any stage in training or practice. As described by Cook et al.,¹⁵ technology-enhanced simulation was defined as any activities involving manikins, task trainers, virtual reality, or computer-based simulations. Papers were excluded if they pertained to technical surgical competence-assessment only, competence in procedures not relevant to EM practice (e.g., laparoscopic surgery), computer-based virtual patients requiring only standard computer equipment, or utilized human patient actors only. Papers were assigned to one of two categories: 1) practical examples of simulation-based assessment; or 2) reviews, meta-analyses, editorials, guidelines, or consensus papers pertaining to simulation-based assessment.

Full-text reviews of included papers were completed independently by six (KC, NR, RW, TM, AP, and CD) investigators to characterize principles of simulation-based assessment. Two investigators (AKH, TC) compiled these principles and summarized and grouped them using a thematic analysis involving the Consensus Framework for Good Assessment.¹³ The thematic analysis followed the technique described by Arksey and O'Malley.¹⁴ To enhance rigour, all other study

investigators reviewed and edited the groupings to generate a final list by consensus. All study authors then independently reviewed the extracted principles and translated them individually into actionable recommendations. Two study authors (AKH, TC) compiled and summarized the list of recommendations. This list was then revised by all study authors at two meetings to create a draft set of recommendations on the use of simulation-based assessment.

Papers found in the literature search and identified as practical examples of simulation-based assessment were reviewed by the study investigators and assigned to one or more of the derived recommendations.

Stakeholder survey

We distributed the draft recommendations on simulation-based assessment to 81 simulation and education experts across the Canadian EM community, representing all Fellow of the Royal College of Physicians of Canada (FRCPC)-EM training programs. Using a Qualtrics online survey (Provo, UT), we asked respondents to review the recommendations and answer two free-text questions: 1) what changes would you make to the recommendations; and 2) what other recommendations regarding the use of simulation-based assessment in EM should be added? We distributed the survey to three groups of individuals: 1) Canadian EM residency program directors; 2) Canadian EM Simulation-Educators Research Collaborative members (EM-SERC); and 3) EM simulation educators identified as part of a prior study identifying curricular content for EM simulation-based education.¹⁶ The study authors reviewed and collated all responses and created a second draft set of recommendations for SBA.

Consensus conference

We presented the revised recommendations to 60 participants at the CAEP Academic Symposium on Education Scholarship Consensus Conference in Halifax, Nova Scotia, on May 25, 2019. We asked participants to reflect on four vignettes (Appendix A), each designed to highlight the tensions related to simulation-based assessment, to provide participants with a pragmatic context of the use of simulation-based assessment. In small groups, participants then reviewed, discussed, and provided written feedback on the proposed recommendations. Two study authors (AKH, TC) collated,

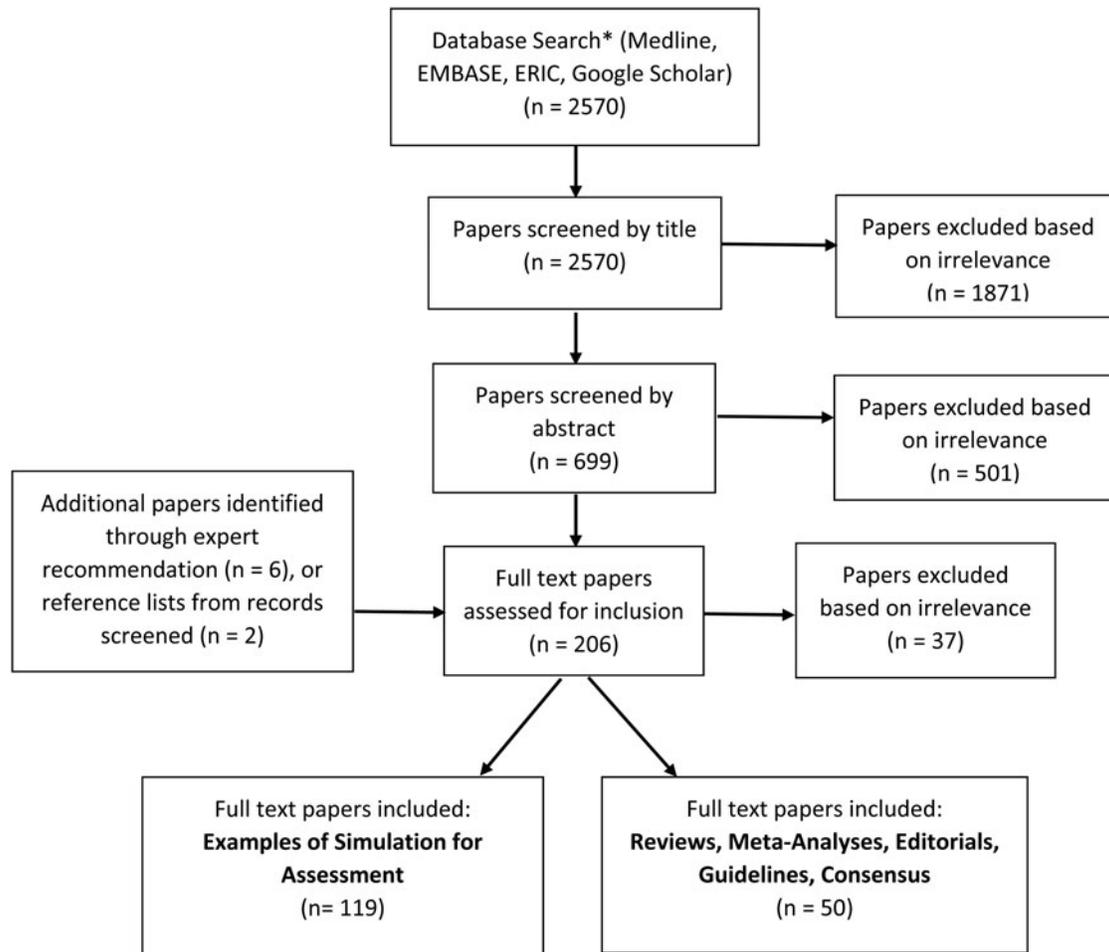


Figure 1. Flow diagram for literature search.

discussed, and incorporated the feedback into the recommendations, generating a final consensus list of recommendations.

RESULTS

Literature review

The literature search is outlined in [Figure 1](#). The database search yielded 2,570 citations after duplicates were removed. Title review excluded 1,871 citations, and abstract review excluded another 501 citations, leaving 198 for full-text review. In addition, six papers were added by study coauthors, and two were found after screening the manuscript reference lists. In total, we performed a full-text review of 206 papers, excluding 37,

allocating 119 to the examples of simulation for assessment category, and the remaining 50 to the reviews, meta-analyses, editorials, guidelines, and consensus papers category.

After analysis of the reviews, meta-analyses, editorials, guidelines, and consensus papers, 209 principles of simulation-based assessment were extracted. These were condensed and summarized by thematic analyses and reviewed by study authors to create a list of 29 principles ([Table 1](#)). Informed by these principles of simulation-based assessment, study authors generated an initial 16 recommendations organized by the Consensus Framework for Good Assessment.¹³ Further, the identified practical examples of simulation-based assessment from the literature search can be found, listed in superscript, after each relevant recommendation ([Table 1](#)).

Table 1. Detailed recommendations for the use of simulation-based assessment in Canadian EM education and training

Framework component	Principles of SBA (Source Reference)	Recommendations for SBA (Recommendation Example)
1. Validity evidence	<p>Assessment tools require generated validity evidence prior to use²⁴⁻²⁹, ideally using a validity framework (i.e., Messick/Kane),^{29,30} with the degree of validity evidence corresponding to the stakes of assessment.^{25,30,31}</p> <p>Despite limited and variable validity evidence that supports the use of simulation in high-stakes assessment,^{24,28-30,32-39} in some cases, this may be a reasonable use of simulation while validity evidence accumulates.⁴⁰⁻⁴²</p> <p>SBA has strongest validity evidence for procedural/technical skill competency,⁴³⁻⁴⁶ crisis resource management, and teamwork.^{28,31,34,36,47,48}</p> <p>Reliability is critical for SBA validity⁴⁹ (see <i>reproducibility</i> below).</p> <p>Generalizability is improved and bias reduced with content blueprinting^{30,37,50,51} and content sampling^{37,39,50,52} as SBA is context and content specific.</p> <p>Appropriate fidelity and authenticity are critical for SBA, with simulator selected to meet the functional properties of the task to be assessed, learner level, task complexity, and stakes.^{33,43,49,50,53-55}</p> <p>Important correlates of SBA include other assessment types,^{29,50} level of trainee or experience (discriminatory validity),^{29,38} and performance in the real world (concurrent validity).^{24,28,30,39,40,42,45,56}</p>	<p>1.1 SBA should prioritize tools and processes with prior validity evidence when possible, with the degree of validity evidence corresponding to the stakes of the assessment.⁵⁷⁻¹⁰⁷</p> <p>1.2 SBA design should align with the functional properties and complexity of the task being assessed, the learner level, and the stakes of assessment.^{58,59,61,83,88-90,105,107-116}</p> <p>1.3 SBA should be considered for medium to higher-stakes assessment but must ensure a standardized, reliable, and reproducible assessment process.^{58,59,61,65,68,82,86,88-90,105,107,110,113,117-124}</p> <p>1.4 EM learners should have adequate exposure to SBA for learning (no or low stakes) prior to SBA of learning (medium or higher stakes) to ensure familiarity and comfort within the simulation environment.^{110,113,125}</p>
2. Reproducibility and equivalence	<p>SBA affords the opportunity to standardize assessment by reducing variability in stimulus more effectively than assessment involving real patients.^{43,52,126,127}</p> <p>Factors that increase SBA reliability include increasing case/content sampling,^{39,50} standardized simulator responses, and confederate role portrayal,³¹ increasing the number of raters,^{51,128} rater training in use of assessment tools,^{24,27,46,49,129} standard setting,^{39,46} and blinded/independent assessors.¹³⁰</p> <p>Checklists are useful for novice learners, though reward thoroughness over skill,^{24,51,53,130,131} and were previously thought to be more reliable than GRS.^{31,50} However, modern GRSs likely have better overall reliability with higher inter-item and inter-station reliability,¹³² are more effective at discriminating expertise,^{31,37} and can be used across multiple tasks.¹³²</p>	<p>2.1 Higher-stakes SBA should include standardization of several key factors (case content and delivery, confederate training, equipment, and rater training) and ensure scenarios are rigorously developed, peer reviewed, and beta tested.^{58,59,61,63,65,86-93,100,112,118,123,124,133-136}</p> <p>2.2 SBA tools should be selected based on the purpose of the assessment and ideally align with clinical assessment tools in use, with preference for GRS-based tools, if possible.^{61,62,65-68,82,85-91,93,135-139}</p> <p>2.3 Facilitators of higher-stakes SBA should receive training, have experience with SBA, and receive periodic rater calibration.^{57,63,65,87,97,105,120,134,140}</p>

(Continued)

Table 1. Continued.

Framework component	Principles of SBA ^(Source Reference)	Recommendations for SBA ^(Recommendation Example)
3. Feasibility	<p>SBA is more reliable for assessment of overall performance than assessment of individual competencies.¹²⁸</p> <p>Reproducibility across institutions can be increased with standardized SBA scenarios and assessment tools across institutions and programs,²⁸ while factoring in variable site technology/equipment.⁵⁴</p> <p>SBA is more resource intensive than many other assessment methods.^{46,128,141-143} Partial-task trainers and low-fidelity equipment can be used in place of high-fidelity manikins for certain assessments, depending on stakes and content.³³</p> <p>While many assessment tools exist with evidence of validity, many are content/context specific, and there is an overall lack of tools with validity evidence for the breadth of contexts/competencies.³⁴</p>	<p>3.1 SBA is a relatively high resource educational modality, and universities should ensure that the local resources required to appropriately conduct SBA are in place before implementation.^{90,114,116,122,144-147}</p> <p>3.2 SBA can be used with a spectrum of simulators and experiences (including low to high fidelity), selected based on the stakes and objective of assessment, as well as the institution's resources.^{58,68,83,97,111,133,134,145-147}</p> <p>3.3 Training programs should share SBA resources, such as cases and assessment tools, whenever possible.^{94,97}</p>
4. Education and catalytic effects	<p>SBA affords the opportunity to deliver targeted feedback in a controlled environment^{25,33,142}, and it is this debriefing that is the most critical component of SBA to maximize the catalytic effect¹⁴¹</p> <p>SBA can inform self-assessment and practice improvement^{25,33}</p> <p>SBA can identify latent errors within teams or systems, resulting in institutional catalytic effect³⁷</p>	<p>4.1 SBA should include feedback to participants as it represents the most critical component of simulation-based learning; this may be adjusted to fit within the type of assessment, with assessment <i>for learning</i> requiring feedback, while assessment <i>of learning</i> should consider feedback provision, if possible.^{63,88,89,97,114,125,134,148-150}</p> <p>4.2 SBA should be utilized for its educational effect (motivating preparation) on the learner and its catalytic effect (feedback from assessment) on both the learner and the education system at large.^{121,125,136,148,149,151-154}</p>
5. Acceptability	<p>SBA can afford the opportunity to perform assessment in a safe environment without harm to real patients.^{37,40,42,127,142,155,156}</p> <p>The cost of simulation can be prohibitive.¹⁴²</p> <p>Learner and clinician "buy-in" is critical for the implementation of SBA.¹⁵⁷</p> <p>There is variability/mixed-opinions about the utility of SBA^{44,155} and variable comfort in the simulation environment.³⁷</p> <p>SBA is generally more acceptable for formative (low stakes) than summative (high stakes) assessment.³⁴</p> <p>SBA has increasing acceptability, having been incorporated into several accreditation and licensure processes (e.g., anesthesia)¹⁵⁸ and the anticipated increasing use of technology-enhanced assessment in the future.¹⁵⁹</p>	<p>5.1 While there is increasing acceptability of SBA, stakeholders such as faculty, learners, program administrators, and national certifying organizations should be consulted both prior to its implementation and, concurrently, to inform program evaluation and encourage acceptability.^{58,92,160-162}</p> <p>5.2 To protect the "safe space" of simulation and encourage acceptability of SBA, the stakes of assessment should always be explicit, and the use of SBA should be balanced by adequate "no stakes" opportunities for learning and practice.</p>

(Continued)

Table 1. Continued.		
Framework component	Principles of SBA (Source Reference)	Recommendations for SBA (Recommendation Example)
6. Part of a program of assessment	<p>Simulation affords the opportunity to assess at the “shows-how” level in Miller’s pyramid,^{26,27,36,53} assessing clinical skills/competencies that are not amenable to assessment in the real world (low frequency, high stakes)^{30,40,43,53,163} or other assessment modalities.^{33,34}</p> <p>Simulation is not ideal for the assessment of knowledge, more appropriate for observable abilities.⁵⁰</p> <p>SBA can be used to assess constituent parts of a procedure/skill prior to real-world performance.⁵⁶</p> <p>Curriculum and assessment committees who include members with SBA expertise will facilitate inclusions/implementation of SBA in programs of assessment.⁵⁴</p>	<p>6.1 SBA should be part of a robust program of assessment, specifically to assess aspects of competence (e.g., Entrustable Professional Activities) that are not more easily and efficiently assessed in the workplace or using other assessment modalities.^{89,94,97,123,125,164-169}</p> <p>6.2 Curriculum and assessment committees should consult or include members with SBA expertise to implement SBA into programs of assessment.</p> <p>6.3 SBA should be considered for use to demonstrate procedural competence prior to attempting certain high-risk procedures in the workplace.^{59,106,107,110,111,122,138,151,152,165,170-174}</p>
<p>EM = emergency medicine; GRS = Global Rating Scales; SBA = simulation-based assessment. References found in Appendix B.</p>		

Stakeholder survey

Of the 81 invited stakeholders, 33 (41%) responded, representing 12 of the 14 Canadian FRCPC-EM training programs. Fifteen respondents identified themselves as simulation educators, 13 identified themselves as program directors, and 5 did not identify their educational role. The recommendations were revised based on survey feedback, and one additional recommendation was incorporated.

Summary of recommendations

We collated and summarized the feedback from the 60 participants at the 2019 CAEP Academic Symposium on Education Scholarship Consensus Conference. The final 6 summary recommendations and 17 detailed recommendations for simulation-based assessment are listed in Tables 1 and 2. The papers referenced in Table 1 can be found in Appendix B.

DISCUSSION

Using a comprehensive process for a literature review, national multi-level stakeholder consultation, and consensus generation, we have developed a set of recommendations for the use of simulation for assessment.

Grounded in established principles of good assessment,¹³ these recommendations are intended to be pragmatic and feasible for frontline educators.

How to use this report

This report is important for educators, program directors, and simulation leaders as they seek to implement simulation-based assessment and collectively navigate this transformative time for postgraduate medical education. Deconstruction of the principles of assessment makes this seemingly large undertaking more manageable and accessible. These recommendations can be used as both a guide to ensure the effective implementation and use of simulation-based assessment, as well as a tool to promote and advocate for high-quality simulation-based assessment. The utility of this report will vary depending on the institution and their existing simulation infrastructure and assessment processes. For some, this may be used as a reference tool to review programs of assessment systematically that already include simulation-based assessment. For others, this may be used to advocate for resources in order to establish a program of assessment that leverages the potential of simulation-based assessment, enabling assessment of performance in domains that are otherwise difficult to

Table 2. Summary recommendations for the use of simulation-based assessment in Canadian EM education and training

1. Validity evidence for assessment tools and processes in SBA should be aligned with the learner level and stakes of assessment.
2. SBA processes such as rater training, case content, and assessment tools should be standardized to support the reproducibility of assessment.
3. SBA is resource intensive, so educators should utilize it only when other assessments will be less effective and match the level of fidelity to the objectives of assessment to minimize cost.
4. When performing SBA, educators should consider its educational effects and provide feedback to participants.
5. When designing SBAs, educators should engage in regular program evaluation and stakeholder consultation to ensure acceptability.
6. Educators should thoughtfully and purposefully incorporate SBA as part of a robust program of assessment

SBA = simulation-based assessment.

capture and illuminating the “blind spots” in current programs of assessment.

Notable findings

As with any assessment method, it is important to question the validity of the assessment process and the decisions that are made based on that assessment (e.g., pass/fail). As our literature review revealed, there is a strong and growing body of validity evidence supporting the use of simulation-based assessment such that we can proceed comfortably using simulation for both lower-stakes assessment and the assessment of specific aspects of competence such as procedural competence and crisis resource management.¹⁵ Simulation-based assessment is likely best used for the assessment of observable skills, attitudes, and behaviours and is less suited for the assessment of knowledge. Ideally, educators should seek out tools and processes that have been evaluated, capitalizing on work that has already been done and adapting to local contexts. As we increase the stakes of the assessment, the demand increases for validity evidence for the assessment process.¹⁷ We recommend that EM programs in Canada start using simulation-based assessment initially with tools and processes that have some form of validity evidence when possible. We also challenge the EM community to continue developing validity evidence for more simulation-based

assessment tools and programs and, in particular, for competencies that are not otherwise easily assessed.

One of the main advantages of simulation is its inherent potential for standardization; the assessment environment can be controlled to deliver the same “stimulus” to all learners. This gives simulation-based assessment a powerful and unique advantage over traditional workplace-based assessment. However, the challenges in delivering such standardized assessments cannot be overstated. Simulation-based assessment is resource intensive, as compared with most existing methods of assessment. With a need to focus on standardization of scenario delivery, training of confederates, and rater training, the costs and efforts to deliver high-quality assessment can become prohibitive. It is important to remember that high-fidelity simulation equipment is not required to conduct simulation-based assessment. Depending on the assessment objectives, task trainers and low-fidelity equipment may be the optimal choices to achieve functional task alignment.¹⁸ Simulation-based assessment resources such as cases or assessment tools should be shared between programs whenever possible to improve access to the highest quality simulation content.

While the benefits of using simulation for medical education are well articulated,^{9,19} the educational effect (the learning that takes place before the assessment) and catalytic effect (the learning that is stimulated by the assessment) of simulation-based assessment have not been well defined. Effective assessment should encourage both of these and promote the learner’s progression to competence.²⁰ Literature suggests that much of the learning that occurs in simulation is a direct result of timely, objective, and constructive feedback from both peers and faculty involved in the debriefing process.²¹ Therefore, we must ensure that processes of simulation-based assessment do not become hurdles to overcome but rather opportunities for concrete feedback, reflection, and enhanced performance.²²

Simulation-based assessment will not succeed if the training programs, content creators, assessors, and participants are not involved in its planning, implementation, and review. If trainees no longer look forward to learning in the simulation lab because of a poorly designed simulation-based assessment, we will lose the learning environment that we have worked so hard to create. The simulation environment is artificial and requires experience for learners to become comfortable, thus making frequent exposure in low-stakes contexts critical if we

are to also use it for assessment. Further defining the stakes of any simulation-based activity can help set the stage for trainee comfort with a shared understanding of expectations and clarity of purpose.²³

LIMITATIONS

While the scope of the literature search was broad, articles retrieved were limited to the English language. Additionally, given the number of articles to be reviewed, each paper was read in detail by a single reviewer; thus, we cannot rule out that some subjectivity may exist in the interpretation and extraction of the key recommendations by individual reviewers who screened the papers. Lastly, the structure of the key recommendations aligns with the 2018 Consensus Framework for Good Assessment¹³; however, we acknowledge that there are other assessment frameworks in existence that may offer different insights into the use of simulation-based assessment.

CONCLUSION

We have proposed a set of recommendations that may serve as a guide in the implementation of simulation-based assessment, as well as advocate for its use with competency-based medical education. While debates regarding simulation-based assessment will (and should) continue until simulation-based assessment has found its thread in the fabric of competency-based medical education, these consensus recommendations, grounded in literature and expert review, should add to the momentum of this important assessment modality.

Supplemental material: The supplemental material for this article can be found at <https://doi.org/10.1017/cem.2019.488>.

Acknowledgements: The study authors would like to thank Jeff Perry as the chair of the CAEP Academic Section, Ian Stiell as the *CJEM* Editor-in-Chief, and all participants who attended the 2019 CAEP Academic Symposium on Education.

Competing interests: This work was supported by the CAEP Academic Section.

REFERENCES

- Holmboe ES, Sherbino J, Englander R, et al. A call to action: the controversy of and rationale for competency-based medical education. *Med Teach* 2017;39(6):574–81.
- Frank JR, Snell LS, Sherbino J. CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada; 2015.
- Harris P, Bhanji F, Topps M, et al. Evolving concepts of assessment in a competency-based world. *Med Teach* 2017;39(6):603–8.
- Wang EE, Quinones J, Fitch MT, et al. Developing technical expertise in emergency medicine—the role of simulation in procedural skill acquisition. *Acad Emerg Med* 2008; 15(11):1046–57.
- Griswold S, Frallicardi A, Boulet J, et al. Simulation-based education to ensure provider competency within the health care system. *Acad Emerg Med* 2018;25(2):168–76.
- Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med* 2003; 78(8):783–8.
- Chiu M, Tarshis J, Antoniou A, 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(12):1357–63.
- Russell E, Hall AK, Hagel C, et al. Simulation in Canadian postgraduate emergency medicine training - a national survey. *CJEM* 2018;20(1):132–41.
- McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ* 2014;48(4):375–85.
- Brydges R, Hatala R, Zendejas B, Erwin PJ, Cook DA. Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Acad Med* 2015;90(2):246–56.
- 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(1):9.
- Rudolph JW, Raemer DB, Simon R. Establishing a safe container for learning in simulation: the role of the presimulation briefing. *Simul Healthc* 2014;9(6):339–49.
- Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach* 2018; 40(11):1102–9.
- Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8(1):19–32.
- Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med* 2013;88(6):872–83.
- Kester-Greene N, Hall AK, Walsh CM. [Simulation curricular content in postgraduate emergency medicine: A multicentre Delphi study]. *CJEM* 2019;21(5):667–75.
- Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ* 2015;49(6):560–75.
- Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. *Acad Med* 2014;89(3):387–92.
- McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med* 2011;86(6):706–11.

20. Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011; 33(3):206–14.
21. Lateef F. Simulation-based learning: just like the real thing. *J Emerg Trauma Shock* 2010;3(4):348–52.
22. Harrison CJ, Könings KD, Schuwirth L, Wass V, van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ Theory Pract* 2015;20(1):229–45.
23. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ* 2019;53(1):76–85.