

Simulation vs workplace-based assessment in resuscitation: a cross-specialty descriptive analysis and comparison

Évaluation des compétences en réanimation en séance de simulation et en milieu de travail : analyse descriptive et comparative de trois spécialités

Jeremy D Seed,¹ Stephen Gauthier,² Boris Zevin,³ Andrew K Hall,⁴ Timothy Chaplin¹

¹Department of Emergency Medicine, Queen's University, Ontario, Canada; ²Department of Medicine, Queen's University, Ontario, Canada;

³Department of Surgery, Queen's University, Ontario, Canada; ⁴Department of Emergency Medicine, University of Ottawa, Ontario, Canada

Correspondence to: Timothy Chaplin, 76 Stuart Street, Kingston, Ontario K7L 2V7; fax: 613 548-1374; phone: 613 548-2368

Published ahead of issue: Feb 20, 2023; CMEJ 2023 Available at <https://doi.org/10.36834/cmej.73692>

© 2023 Seed, Gauthier, Zevin, Hall, Chaplin; licensee Synergies Partners. This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License. (<https://creativecommons.org/licenses/by-nc-nd/4.0>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited.

Abstract

Background: Simulation-based assessment can complement workplace-based assessment of rare or difficult to assess Entrustable Professional Activities (EPAs). We aimed to compare the use of simulation-based assessment for resuscitation-focused EPAs in three postgraduate medical training programs and describe faculty perceptions of simulation-based assessment.

Methods: EPA assessment scores and setting (simulation or workplace) were extracted from 2017-2020 for internal medicine, emergency medicine, and surgical foundations residents at the transition to discipline and foundations of discipline stages. A questionnaire was distributed to clinical competency committee members.

Results: Eleven percent of EPA assessments were simulation-based. The proportion of simulation-based assessment did not differ between programs but differed between transition (38%) and foundations (4%) stages within surgical foundations only. Entrustment scores differed between settings in emergency medicine at the transition level only (simulation: 4.82 ± 0.60 workplace: 3.74 ± 0.93). 70% of committee members (n=20) completed the questionnaire. Of those that use simulation-based assessment, 45% interpret them differently than workplace-based assessments. 73% and 100% trust simulation for high-stakes and low-stakes assessment, respectively.

Conclusions: The proportion of simulation-based assessment for resuscitation focused EPAs did not differ between three postgraduate medical training programs. Interpretation of simulation-based assessment data between committee members was inconsistent. All respondents trust simulation-based assessment for low-stakes, and the majority for high-stakes assessment. These findings have practical implications for the integration simulation into programs of assessment.

Résumé

Contexte : Pour les activités professionnelles confiées (APC) qui sont rarement observées ou difficiles à évaluer, l'évaluation dans les séances de simulation peut compléter l'évaluation en milieu de travail. Nous avons comparé l'utilisation de l'évaluation lors de simulations pour les APC axées sur la réanimation dans trois programmes de formation médicale postdoctorale et décrit les perceptions de membres du corps professoral concernant cette modalité d'évaluation.

Méthodes : Nous avons extrait les scores et le cadre (simulation ou lieu de travail) d'évaluation des APC de 2017 à 2020 pour les résidents en médecine interne, en médecine d'urgence et en fondements chirurgicaux aux étapes de transition vers la discipline et de fondements de la discipline. Un questionnaire a été distribué aux membres du comité des compétences cliniques.

Résultats : Onze pour cent des évaluations d'APC étaient faites lors de séances de simulation. Cette proportion était la même pour tous les programmes, mais dans le cadre des fondements chirurgicaux, elle était différente selon qu'il s'agissait de l'étape de transition (38 %) ou de l'étape des fondements (4 %). Les scores de confiance différaient selon le cadre de l'évaluation uniquement pour les résidents en médecine d'urgence à l'étape de la transition (simulation : $4,82 \pm 0,60$; lieu de travail : $3,74 \pm 0,93$). Le questionnaire a été rempli par 70 % des membres du comité (n=20). Parmi ceux qui avaient utilisé l'évaluation en séance de simulation, 45 % avaient interprété les données de l'évaluation différemment de la façon dont ils interprètent les données d'évaluation en milieu de travail. Soixante-treize pour cent et 100 % d'entre eux font confiance à la simulation pour les évaluations à enjeux élevés et à faibles enjeux, respectivement.

Conclusions : La proportion d'évaluations en séance de simulation pour les APC axées sur la réanimation était la même pour trois programmes de formation médicale postdoctorale. Les membres du comité n'ont pas interprété les données de ce type d'évaluation de manière uniforme. Tous les répondants font confiance à l'évaluation en séance de simulation pour les évaluations à faible enjeu, et la plupart d'entre eux pour les évaluations à enjeu élevé. Ces résultats ont des incidences pratiques sur l'intégration de la simulation dans les programmes d'évaluation.

Introduction

The Royal College of Physicians and Surgeons of Canada (RCPSC) has implemented a version of competency-based medical education (CBME), termed Competence by Design (CBD), across all Canadian specialty postgraduate training programs.¹ The program of assessment in CBD is based upon specialty-specific entrustable professional activities (EPAs) organized into four stages of training,² with priority placed on frequent formative assessments based on direct observation of a learner's abilities.³ In the first stage, transition to discipline, new postgraduate medical trainees are oriented to residency education and assessed for readiness to work as a learner within the discipline. In the second stage, foundations of discipline, trainees are taught and assessed on broad-based competencies before progressing to discipline-specific competencies.²

Ideally, the assessment of EPAs should be performed in authentic environments, such as the workplace. However, the workplace-based assessment of resuscitation-focused EPAs that are rare or occur at unpredictable times has proved to be challenging.^{4,5} Simulation allows for frequent and predictable assessment in a safe and reproducible environment, and may complement a workplace-based program of assessment for certain EPAs.⁶ The validity evidence for simulation-based assessment is growing, however the "extrapolation inference" in Kane's validity framework is often cited as a limitation to its use.^{7,8} This refers to the 'inferential leap' or assumption that performance in the simulation laboratory is a valid representation of performance in the 'real world'. It also remains unclear how training programs in Canada are using simulation-based assessment. To optimize CBME programs of assessment, an understanding of the interpretation and use of simulation-based assessment is required.

We sought to describe the use of simulation-based assessment for resuscitation-focused EPAs in three postgraduate medical training programs. Our secondary objectives were to 1) compare entrustment scores assigned in the simulation setting to those in the workplace and 2) solicit the perspectives of clinical competency committee members related to simulation-based assessment.

Methods

We conducted a retrospective study of postgraduate medical resident assessment data from three RCPSC training programs in addition to collecting survey data at Queen's University in Canada. The Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board approved this study (REB ID#:6030091).

Residency training programs at Queen's University had access to a funded simulation laboratory with technical support for both manikin-based and task-trainer sessions. At the time of this study, the use of simulation varied between programs. Emergency Medicine participated in weekly simulations in the first and second stages. Internal Medicine participated in a single simulation session in the first stage and four in the second stage.⁹ Surgical Foundations used simulation on a weekly basis during a six-week 'bootcamp' in the first stage, and four sessions in the second.⁹ Assessment of EPAs in the simulation lab was optional for all sessions, except during the Surgical Foundations bootcamp.

Queen's University implemented CBD in July 2017, although training programs transitioned variably to entrustment-based assessment. Emergency Medicine used a five-point entrustment scale¹⁰ from July 1, 2017 to July 17, 2018, after which time the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) scale,¹⁰ was adopted. Starting in July 2018, Surgical Foundations used the O-SCORE and Internal Medicine used a six-point entrustment scale¹¹ (Appendix A). The Surgical Foundations program encompasses all surgical programs at Queen's except for ophthalmology. These programs were selected as they represent the majority of residents (38/57, 67%) in RCPSC specialty training programs.

Assessment data

We identified resuscitation-focused EPAs in the first two CBD stages of training across Internal Medicine, Emergency Medicine, and Surgical Foundations (Table 1). Assessment data was extracted from the institution-wide electronic portfolio (Elentra, Kingston, ON) from the implementation of entrustment scoring for each program to June 30, 2020. Assessment data was de-identified prior to analysis. To allow for comparisons within programs to be made, entrustment scores from both five-point scales utilized by Emergency Medicine were considered equivalent for analysis.

Table 1. Description of program specific resuscitation-focused EPAs for the first and second stages of training

| EPA Description | |
|-----------------------|--|
| Emergency Medicine | |
| 1 st Stage | Recognizing the unstable/critically ill patient, mobilizing the health care team and supervisor, and initiating basic life support |
| 2 nd Stage | Initiating and assisting in resuscitation of critically ill patients |
| Internal Medicine | |
| 1 st Stage | Identifying and assessing unstable patients, providing initial management, and obtaining help |
| 2 nd Stage | Assessing unstable patients, providing targeted treatment and consulting as needed |
| Surgical Foundations | |
| 1 st Stage | Recognizing and initiating early management for critically ill surgical patients |
| 2 nd Stage | Providing initial management for critically ill surgical patients |

Questionnaire

We developed a questionnaire using Qualtrics XM (Qualtrics, Provo, Utah) to capture the perspectives of clinical competency committee members (referred to as committee members going forward) on the interpretation and use of simulation-based assessment data (Appendix B). The questionnaire was piloted by Program Directors in Emergency Medicine, Internal Medicine, Surgical Foundations and two clinician-educators to ensure clear language and was revised based on feedback. The questionnaire was distributed via electronic-mail to all 20 committee members in Emergency Medicine, Internal Medicine, and Surgical Foundations. Responses were anonymous and free text answers were collated and summarized.

Analysis

Data analysis was completed using SPSS Statistics Version 26.0 (IBM Corp). Descriptive statistics included the number of EPA assessments within each program, assessment setting and stage of training. The proportion of assessments in the simulation and workplace settings were compared between programs, stage of training, and academic year using a Chi-Square Test. Entrustment scores were compared using an independent samples t-test. Statistical significance was considered to be $p < 0.05$ and all values are presented as mean \pm SD.

Results

Use of simulation-based assessment

The research team extracted a total of 682 resuscitation-focused EPA assessments. Of these, 75 were simulation-based. The proportion of simulation-based assessment (Table 2) did not differ between programs ($p = 0.28$). A greater proportion of simulation-based assessment was completed in the first stage (51/270) compared to the second stage (24/412; $p < 0.001$). The proportion of simulation-based assessment since the implementation of CBD was only different in Emergency Medicine, increasing from the 2017-2018 (2/48; $p = 0.01$) and 2018-2019 (3/94; $p < 0.001$) academic years to 2019-2020 (20/98).

Assessment setting

There was a difference in entrustment scores between settings (Table 2) in the first stage of training ($p < 0.001$), but not the second stage ($p = 0.22$) within the Emergency Medicine program. Entrustment scores did not differ significantly between clinical settings in any other program at any stage of training.

Table 2. Number of assessments as n (%) and entrustment scores as mean (SD) broken down by program, clinical setting, and program specific EPA.

| | Number of Assessments (%) | | Entrustment Scores (SD) | |
|-----------------------|---------------------------|------------|-------------------------|--------------|
| | Workplace | Simulation | Workplace | Simulation |
| 5-Point Scale | | | | |
| Emergency Medicine | 215 (90) | 25 (10) | | |
| 1 st Stage | 73 (87) | 11 (13) | 3.74 (0.94)* | 4.82 (0.60)* |
| 2 nd Stage | 142 (91) | 14 (9) | 3.72 (0.83) | 4.00 (0.68) |
| Surgical Foundations | 238 (87) | 36 (13) | | |
| 1 st Stage | 48 (62) | 29 (38)* | 4.21 (0.77) | 4.10 (0.90) |
| 2 nd Stage | 190 (96) | 7 (4)* | 4.07 (0.87) | 4.00 (0.58) |
| 6-Point Scale | | | | |
| Internal Medicine | 154 (92) | 14 (8) | | |
| 1 st Stage | 98 (90) | 11 (10) | 4.15 (0.87) | 4.00 (1.00) |
| 2 nd Stage | 56 (95) | 3 (5) | 4.54 (0.71) | 3.67 (1.16) |

* $p < 0.05$

Questionnaire

Fourteen of 20 committee members completed the questionnaire. Eleven of 14 reported their programs use simulation-based assessment. Of these 11, five interpret assessment data differently depending on the setting, citing a higher perceived stress level and degree of complexity in the workplace as well as the ability to provide more focused observation in the simulation setting. The other six of 11 interpret simulation-based assessment and

workplace assessment data, similarly, reporting both methods require similar skills, provide valid assessment, and cover similar clinical scenarios. Of the 11 respondents who use simulation-based assessment, nine believed that simulation-based assessment should be used more often, eight would trust it for high-stakes, and all 11 for low-stakes assessment.

Discussion

We have described the use of simulation-based assessment of resuscitation-focused EPAs across three PGME training programs at one academic institution in Canada. Overall, simulation accounted for 11% of resuscitation-focused EPA assessments and the proportion of simulation-based assessments did not differ significantly between programs. This was unanticipated as the use of simulation for resuscitation skills training within each training program differed substantially (i.e. weekly vs a single session). This suggests that training and assessment are discrete objectives and should be explicitly planned into programs of assessment for resuscitation focused EPAs.

The proportion of simulation-based assessment at the first and second CBD stages of training were similar within Emergency Medicine and Internal Medicine, but not the Surgical Foundations training program (38% first stage vs 4% second stage). This differential use is related to the requirement of simulation-based assessment during the six-week Surgical Foundations 'bootcamp' in the first stage, as well as a shift from a resuscitation focus to specialty-specific surgical training in the second stage. The use of simulation-based assessment since the implementation of CBD was seen to increase only in the Emergency Medicine training program. This reflects the current dialogue of an increasing trend in simulation-based assessment in response to challenges with workplace-based assessment. For example, although specific to Emergency Medicine, recent studies found that only two programs utilized simulation for assessment in 2018 (prior to CBD) and residents were not meeting the recommended number of EPA assessments in the first stage of training early in the implementation of CBD.^{5,12}

Overall we found no significant difference between simulation-based and workplace-based entrustment scores. Across all three programs, using program specific entrustment-based assessment tools, simulation-based assessment scores were only higher for a first stage Emergency Medicine EPA. The results from three studies correlating resuscitation performance in the workplace and

simulation laboratory range from no correlation to a moderate positive correlation.¹³⁻¹⁵ The most recent study found that on average workplace-based entrustment scores were higher but there was no correlation at the individual resident level.¹³ Our finding that entrustment scores did not differ between the workplace and simulation settings across three training programs and two stages adds to the increasing evidence to support the use of simulation-based assessment of resuscitation-focused EPAs.

Of the competence committee members whose programs use simulation for assessment, the majority trusted its use for both low and high-stakes assessment and all reported that it should be used more frequently. However, not all interpreted simulation-based assessment data in the same way due to the perceived differences in the stress, complexity, and nature of observation in the simulated setting. This finding reflects the ongoing perceived gap that exists between the simulated and clinical settings as described in Kane's validity framework.⁸ In order to fully integrate simulation-based assessment into programs of assessment, future efforts are required to bridge this perceived gap^{7,16} as acceptability amongst stakeholders is key to establishing a good system of assessment.⁴

The current study has limitations. First, the research team did not have access to individual resident data, so correlational analysis was not possible. Second, we did not analyze narrative assessment data and the quality of feedback may vary between assessment settings.

Conclusions

The expectation that all EPA assessments are workplace-based, especially for rare or time-sensitive cases, is unrealistic. We found that simulation was used for approximately 11% of resuscitation-focused EPA assessments and that entrustment scores did not differ substantially between the workplace and simulation setting, except for one Emergency Medicine EPA. Importantly, we have also found that the majority of committee members already using simulation-based assessment trust it to aid in their progression decisions and believe that it should be used more often.

Conflicts of Interest: There are no conflicts of interest declared by the authors, financial or otherwise.

Funding: There was no funding for this project.

References

- Busing N, Harris K, MacLellan AM, et al. The future of postgraduate medical education in Canada. *Acad Med J Assoc Am Med Coll.* 2015;90(9):1258-1263. <https://doi.org/10.1097/ACM.0000000000000815>
- Frank JR, Snell LS, Sherbino J. *CanMEDS 2015 Physician Competency Framework.* 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):603-608. <https://doi.org/10.1080/0142159X.2017.1315071>
- Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach.* 2018;40(11):1102-1109. <https://doi.org/10.1080/0142159X.2018.1500016>
- Thoma B, Hall AK, Clark K, et al. Evaluation of a national competency-based assessment system in emergency medicine: a CanDREAM study. *J Grad Med Educ.* 2020;12(4):425-434. <https://doi.org/10.4300/JGME-D-19-00803.1>
- Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, Collaborators for the IC. The role of assessment in competency-based medical education. *Med Teach.* 2010;32(8):676-682. <https://doi.org/10.3109/0142159X.2010.500704>
- 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):872-883. <https://doi.org/10.1097/ACM.0b013e31828ffdcf>
- 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):560-575. <https://doi.org/10.1111/medu.12678>
- McMurray L, Hall AK, Rich J, Merchant S, Chaplin T. The nightmares course: a longitudinal, multidisciplinary, simulation-based curriculum to train and assess resident competence in resuscitation. *J Grad Med Educ.* 2017;9(4):503-508. <https://doi.org/10.4300/JGME-D-16-00462.1>
- Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med J Assoc Am Med Coll.* 2012;87(10):1401-1407. <https://doi.org/10.1097/ACM.0b013e3182677805>
- Peters H, Holzhausen Y, Boscardin C, Cate O ten, Chen HC. Twelve tips for the implementation of EPAs for assessment and entrustment decisions. *Med Teach.* 2017;39(8):802-807. <https://doi.org/10.1080/0142159X.2017.1331031>
- Russell E, Hall AK, Hagel C, Petrosoniak A, Dagnone JD, Howes D. Simulation in Canadian postgraduate emergency medicine training – a national survey. *Can J Emerg Med.* 2018;20(1):132-141. <https://doi.org/10.1017/cem.2017.24>
- Prudhomme N, O'Brien M, McConnell MM, Dudek N, Cheung WJ. Relationship between ratings of performance in the simulated and workplace environments among emergency medicine residents. *Can J Emerg Med.* Published online 2020:1-8. <https://doi.org/10.1017/cem.2020.388>
- 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.* 2019;4. <https://doi.org/10.1186/s41077-019-0099-4>
- Hall AK, Damon Dagnone J, Moore S, et al. Comparison of simulation-based resuscitation performance assessments with in-training evaluation reports in emergency medicine residents: a canadian multicenter study. *AEM Educ Train.* 2017;1(4):293-300. <https://doi.org/10.1002/aet2.10055>
- Calhoun AW, Bhanji F, Sherbino J, Hatala R. Simulation for high-stakes assessment in pediatric emergency medicine. *Clin Pediatr Emerg Med.* 2016;17(3):212-223. <https://doi.org/10.1016/j.cpem.2016.05.001>

Appendix A. Entrustment scales

| Entrustment Scale | Entrustment Scoring |
|-------------------------------------|--|
| Ten Cate (EM 01/07/2017-17/07/2018) | |
| | Observation only |
| | Direct, proactive supervision |
| | Indirect, reactive supervision |
| | Independent performance with remote supervision |
| | Supervision of trainees |
| O-SCORE (SF, EM 18/07/2018-Present) | |
| | I had to do |
| | I had to talk them through |
| | I had to prompt them from time to time |
| | I needed to be in the room just in case |
| | I did not need to be there |
| 6-Point (IM) | |
| | Supervisor actively performs EPA with resident |
| | Supervisor intermittently assists resident to perform the EPA |
| | Supervisor outside the room, immediately available, double-checks findings |
| | Supervisor outside room, immediately available, checks only key findings |
| | Supervisor offsite, available by phone, checks only key findings |
| | Distant supervisor, post-hoc debrief available as needed |

Published ahead of issue

Appendix B. Questionnaire.

Page 2 – Demographic Questions

- 1) Program
- 2) How long have you been part of your program's competency committee?
- 3) Over an academic year, how often do you teach in the simulation lab?
 - a. **Never, 1-2 sessions, 3-5, >5 sessions**

Page 3 Perception of sim for assessment

- 1) Is your program currently using simulation for the assessment of EPAs? **(YES/NO)**

IF YES:

- 1) When reviewing a resident's EPA performance at your competency committee meetings to decide on progression or promotion:
 - a. Has your committee ever considered assessment data obtained from simulated encounters differently than assessment data from real-life clinical encounters? **(YES/NO)**
 - b. Have you ever interpreted assessment data differently based on whether it was from a simulated encounter compared to a real-life situation? **(YES/NO)**
 - c. Do you trust assessment data that comes from the simulated setting more or less than data that comes from the clinical setting? **(More/Less/Same)**
 - i. **Please explain**
 - d. Do you think assessments that come from simulated settings should always be interpreted in the same way as assessments that come from the clinical setting? **(YES/NO)**
 - i. **Please explain**

IF NO:

- a. Would you interpret assessment data differently based on whether it was from a simulated encounter compared to a real-life situation? **(YES/NO)**
 - a. **Please explain**
- b. Would you trust assessment data that comes from the simulated setting more or less than data that comes from the clinical setting? **(More/Less/Same)**
 - ii. **Please explain**

Regardless of whether or not your program is currently using simulation for assessment:

- 1) Please provide your thoughts on the following examples:
 - a. For a given EPA, resident 'A' has received very high entrustment scores in the simulated setting, but low scores in the clinical setting.
 1. **Comments:**
 - b. For a given EPA, resident 'A' has received several assessments in the simulated setting but none in the clinical setting
 1. **Comments:**
- 2) Are there certain EPAs that are better assessed in the simulated setting? **(YES/NO)**
 - a. **Please explain**
- 3) Would you trust "high stakes" (i.e. advancement decision or licensing) assessment performed in the simulation setting?
 - a. **Why or why not?**
- 4) Would you trust "low stakes" (i.e. as part of a program of assessment) assessment performed in the simulation setting?
 - a. **Why or why not?**
- 5) In your department, are the faculty who assess in the simulation setting representative of the whole faculty? **(YES/NO)**
 - a. If not, how are they different when it comes to assessment?
- 6) Should simulation be used more or less frequently for assessment in your program? **(LESS/MORE)**
 - a. **Please explain**
- 7) Any other comments about the use of simulation for assessment?

Thank you