Can J Anesth/J Can Anesth (2021) 68:235–244
https://doi.org/10.1007/s12630-020-01850-x

REPORTS OF ORIGINAL INVESTIGATIONS

Resident competencies before and after short intensive care unit rotations: a multicentre pilot observational study

Compétences des résidents avant et après de courts stages à l'unité de soins intensifs : une étude pilote observationnelle multicentrique

Dominique Piquette, MD, PhD, FRCPC · Alberto Goffi, MD · Christie Lee, MD, MSc, FRCPC · Ryan Brydges, PhD · Catharine M. Walsh, MD, PhD, FRCPC · Briseida Mema, MD, MEd, FRCPC · Chris Parshuram, MBChB, DPhil, FRACP

Received: 26 February 2020 / Revised: 19 August 2020 / Accepted: 22 August 2020 / Published online: 10 November 2020

Abstract

Purpose *Residency programs need to understand the competencies developed by residents during an intensive care unit (ICU) rotation, so that curricula and assessments maximize residents' learning. The primary study objective was to evaluate the feasibility for training programs and acceptability by residents of conducting a multi-competency assessment during a four-week ICU rotation.*

Methods *We conducted a prospective, multicentre observational pilot study in three ICUs. During weeks 1 and 4 of an ICU rotation, we conducted repeated standardized assessments of non-critical care specialty residents' competencies in cognitive reasoning (script concordance test [SCT]), procedural skills (objective structured assessment of technical skills [OSATS]-global rating scale], and communication skills through a written test, two procedural simulations, and a simulated encounter with a "family member". The feasibility outcomes included program costs, the proportion of enrolled residents able to complete at least one three-station assessment during their four-week ICU rotation, and acceptability of the assessment for the trainees.*

Results *We enrolled 63 (69%) of 91 eligible residents, with 58 (92%) completing at least one assessment. The total cost to conduct 90 assessments was CAD 33,800. The majority of participants agreed that the assessment was fair and that it measured important clinical abilities. For the 32 residents who completed two assessments, the mean (standard deviation) cognitive reasoning and procedural skill scores increased between weeks 1 and 4 [SCT difference, 3.1 (6.5), $P = 0.01$; OSATS difference for bag-mask ventilation and central line insertion, 0.4 (0.5) and 0.6 (0.8), respectively; both $P \leq 0.001$]. Nevertheless, the communication scores did not change significantly.*

Conclusions *A monthly multi-competency assessment for specialty residents rotating in the ICU is likely feasible for most programs with appropriate resources, and generally acceptable for residents. Specialty residents' cognitive reasoning and procedural skills may improve during a four-week ICU rotation, whereas communication skills may not.*

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12630-020-01850-x) contains supplementary material, which is available to authorized users.

DominiquePiquette, MD, PhD, FRCPC (✉)
Sunnybrook Health Sciences Centre, University of Toronto, 2075 Bayview Avenue, Room D108, Toronto, ON M4N3M5, Canada
e-mail: dominique.piquette@sunnybrook.ca

A. Goffi, MD
St. Michael's Hospital, University of Toronto, Toronto, ON, Canada

C. Lee, MD, MSc, FRCPC
Mt. Sinai Hospital, University of Toronto, Toronto, ON, Canada

R. Brydges, PhD
The Wilson Centre, University of Toronto, Toronto, ON, Canada

C. M. Walsh, MD, PhD, FRCPC · B. Mema, MD, MEd, FRCPC · C. Parshuram, MBChB, DPhil, FRACP
The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada

Résumé

Objectif *Afin que les programmes de formation et les évaluations maximisent les apprentissages des résidents, les programmes de résidence doivent comprendre quelles compétences sont développées par les résidents pendant un stage à l'unité de soins intensifs (USI). L'objectif principal de cette étude était d'évaluer la faisabilité pour les programmes de formation et l'acceptabilité par les résidents de réaliser une évaluation multi-compétences pendant un stage de quatre semaines à l'USI.*

Méthode *Nous avons réalisé une étude pilote observationnelle prospective multicentrique dans trois USI. Pendant les semaines 1 et 4 du stage à l'USI, nous avons mené des évaluations standardisées répétées des compétences des résidents non inscrits dans une spécialisation en soins intensifs en matière de raisonnement cognitif (test de concordance de script [SCT]), d'habiletés procédurales (évaluation objective structurée des compétences techniques [OSATS] - échelle d'évaluation globale), et d'habiletés de communication via un examen écrit, deux simulations d'intervention, et une rencontre simulée avec un « membre de la famille ». Les critères de faisabilité comprenaient les coûts du programme d'évaluation, la proportion de résidents inscrits capables de compléter au moins une évaluation en trois stations au cours de leur stage de quatre semaines à l'USI, et l'acceptabilité de l'évaluation par les résidents.*

Résultats *Nous avons recruté 63 (69 %) des 91 résidents éligibles, et 58 (92 %) ont complété au moins une évaluation. Le coût total pour réaliser 90 évaluations était de 33 800 CAD. La majorité des participants étaient d'accord que l'évaluation était équitable et qu'elle mesurait d'importantes habiletés cliniques. Chez les 32 résidents ayant complété deux évaluations, les scores moyens (écart type) en matière de raisonnement cognitif et d'habiletés techniques ont augmenté entre les semaines 1 et 4 : différence au SCT, 3,1 (6,5), P = 0,0; différence à l'OSATS pour la ventilation au masque et l'installation d'une voie centrale, 0,4 (0,5) et 0,6 (0,8), respectivement; tous deux P ≤ 0,001. Toutefois, les scores en matière de communication n'ont pas changé de manière significative.*

Conclusion *Une évaluation multi-compétences mensuelle des résidents en spécialisation faisant un stage à l'USI est probablement réalisable dans la plupart des programmes disposant des ressources nécessaires, et elle est généralement acceptable pour les résidents. Le raisonnement cognitif et les habiletés techniques des résidents pourraient s'améliorer pendant un stage de quatre semaines à l'USI, alors que leurs compétences de communication pourraient demeurer inchangées.*

Keywords medical education · internship and residency · assessment · simulation · critical care

In most medical specialty training programs (e.g., internal medicine, general surgery), postgraduate trainees (residents) are required to complete rotations in the intensive care unit (ICU). By the end of their ICU rotations, residents are expected to have developed important clinical abilities inherent to critical care practice, such as initial stabilization of acutely ill patients, invasive procedures, and end-of-life discussions. Nevertheless, the ICU environment presents many educational challenges, including unpredictable clinical activities, high-risk patients, time pressures, and variable patient populations and medical problems across centres.[1–4] The current educational structure creates additional constraints on residents' ICU learning experience. Short rotations (e.g., four weeks), discontinuity of ICU experience (e.g., rotations distributed over time and ICUs, off-service commitments), competition between residents to access learning opportunities, and inconsistent supervision practices are among the factors possibly affecting learning opportunities.[5–7] Learning outcomes resulting from a typical ICU rotation may therefore be limited or vary significantly across domains and individual residents.

Limited evidence suggests that short clinical ICU exposure may be insufficient for residents to achieve acceptable levels of competence in various domains.[8] For example, Ottestad *et al.* reported both poor and acceptable levels of performance in managing simulated septic patients for specialty residents who had completed an ICU rotation.[9] In another study, bedside skills related to respiratory and circulatory systems were assessed at the end of an ICU rotation and compared among internal medicine residents who had and had not completed a four-hour simulation course.[10] Residents who had received simulation-based training in addition to the usual clinical training significantly outperformed their colleagues. Singer *et al.* also showed that first-year residents who had received simulation training performed better than traditionally trained (clinical exposure only) third-year residents during bedside clinical assessments.[11] The authors concluded that "critical care competency cannot be assumed after clinical intensive care unit rotations".[11] These examples illustrate the need to better assess specialty residents' learning outcomes after ICU training.

The small number of resident assessments traditionally completed during ICU rotations—typically one summative end-of-rotation assessment per resident—has likely contributed to our limited understanding of the learning

occurring during an ICU rotation. Assessment practices are, however, rapidly changing now that many specialty programs are transitioning to competency-based medical education (CBME).[12] With CBME, programs must regularly measure residents' level of achievement for important competencies. Each program identifies clinical activities requiring assessment, based on their expectations of what is being done and learned by residents during a rotation. It is not known whether competencies being assessed (or expected to be assessed) in the ICU align with what residents are realistically able to learn. A better understanding of specialty resident learning outcomes resulting from ICU rotations would therefore greatly help with curricular and assessment planning.

To assess clinical competencies that specialty programs commonly expect residents to learn during a typical ICU rotation, we undertook a pilot study of standardized multi-competency assessment for specialty residents rotating through the ICU. The primary objective was to evaluate the i) feasibility for three different ICU training programs to achieve a high resident completion rate of at least one standardized, multi-competency assessment during a four-

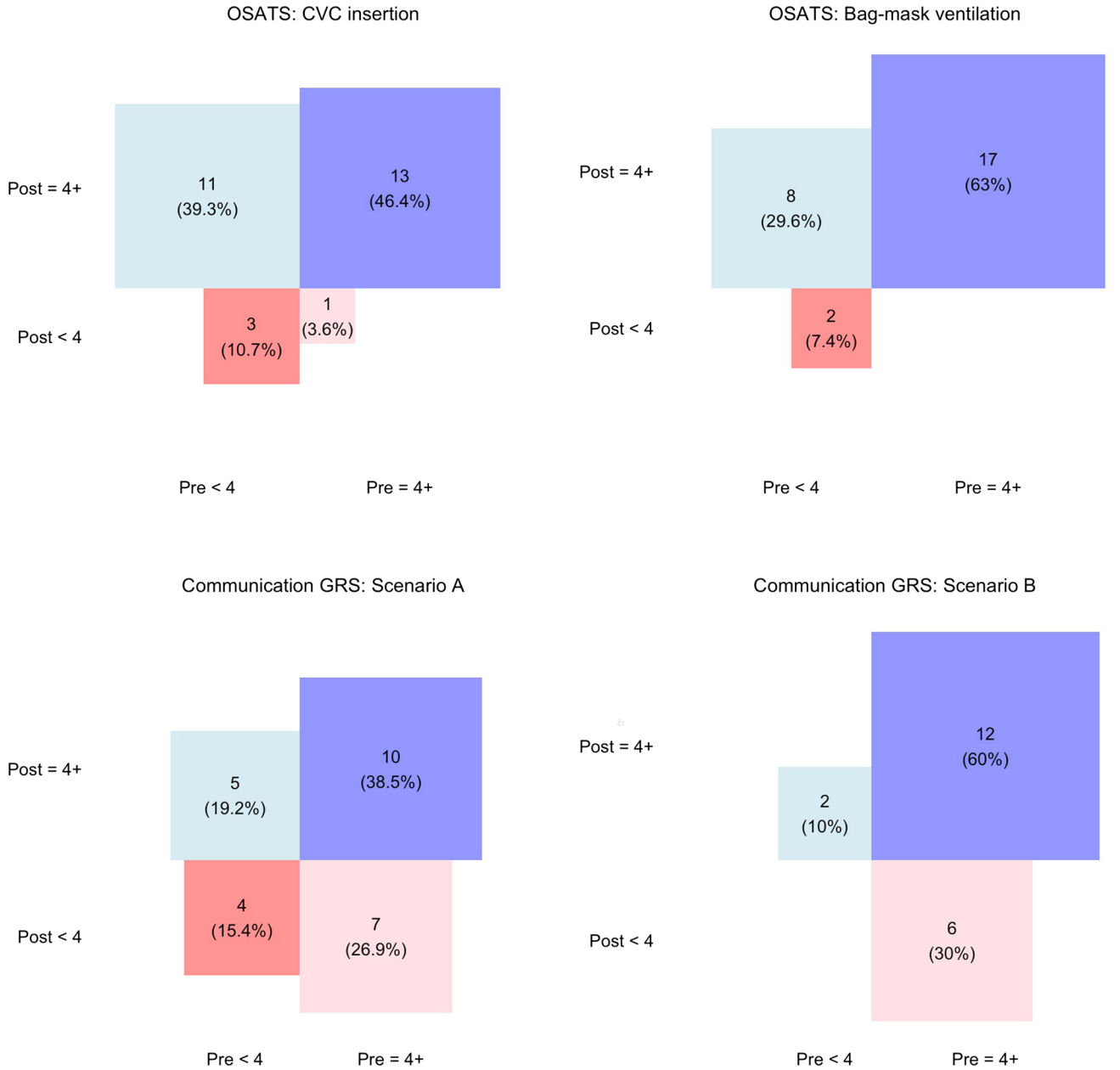

Figure Four-fold plot representing objective structured assessment of technical skills (OSATS) procedural scores and communication global rating scale (GRS) scores for residents who completed the assessment in week 1 and 4

week ICU rotation over consecutive months, ii) assessment costs, and iii) assessment acceptability by residents. The secondary objective was to determine the learning outcomes of non-critical care specialty residents in three domains of competence (cognitive reasoning, procedures, and communication) after a four-week ICU rotation.

Methods

We conducted this study in three mixed medical and surgical ICUs, each located in a different academic medical centre in Toronto, Canada. The study was approved by the Review Ethics Board of each institution.

Residents of any postgraduate year enrolled in a non-critical care, Royal College of Physicians and Surgeons of Canada (RCPSC)-accredited program who were scheduled to rotate in one of the three participating ICUs for at least four weeks were eligible for the study. For study recruitment, we emailed all residents scheduled for an ICU rotation during the study period prior to the beginning of the rotation and met them in person on day one of their ICU rotation. Residents were excluded if they had already participated in the study during a previous ICU rotation (e.g., at another hospital site) at the time of recruitment. Residents provided written informed consent prior to enrolment. Individual follow-up ended after week 4 of the ICU rotation.

Five to ten specialty residents typically rotate monthly at each ICU site. Resident performance is routinely assessed by a mid-rotation and/or end-of-rotation in-training evaluation report based on day-to-day clinical performance. None of the sites used other formal assessments at the time of the study. The transition to the RCPSC Competency by Design model had not affected off-service specialty residents rotating in the ICU at the time of the study.

Standardized multi-competency assessment

To design a standardized multi-competency assessment, we first reviewed common objectives of training targeted by multiple specialties for residents' ICU rotations and conducted a survey of ten local ICU clinician educators and teachers (see eAppendix 1 in the Electronic Supplementary Material [ESM]). Informed by this process, we selected the following three domains of competence: 1) cognitive reasoning (diagnosis, investigation, and management of common ICU medical problems); 2) procedural skills (e.g., central venous catheter [CVC] insertion and bag-mask ventilation [BMV]); and 3) communication skills (providing an initial update post-ICU admission and leading an end-of-life discussion with a simulated family member). Our objective was to independently assess these specific competencies, not to assess overall competence in critical care medicine.

We then conducted an extensive literature review to identify strategies and tools available to assess these domains of competence and weighed each tool according to the existing relevant validity evidence.[13] Assessment tools were selected for each domain of competence. This included the script concordance test (SCT) for cognitive reasoning,[14] objective structured assessment of technical skills-global rating scale (OSATS-GRS) for procedural skills,[15] and a five-item global rating scale for communication skills (see ESM, eAppendices 2, 3, and 4).[16] Next, we developed the three parts of the competency assessment.

Part 1: Written SCT for critical care cognitive reasoning

We developed a SCT measuring clinical reasoning in critical care medicine according to published guidelines.[17–19] We first developed an assessment blueprint based on the assessment priorities in medical expertise established by the ICU educators survey. Two adult ICU physicians then wrote a bank of 70 vignettes to create two versions of a SCT, each initially including 35 clinical vignettes (or cases) and 105 questions (see ESM, eAppendix 2). We decided on the number of vignettes and questions, Likert scale, scoring, panel size, and composition based on existing recommendations.[18] An SCT assesses residents' ability to interpret medical information under conditions of uncertainty.[20,21] For each clinical vignette, a series of questions (items) about diagnoses, investigations, or management are asked. Test optimization included item review by two experienced ICU educators based on a published quality grid[18] and piloting with subspecialty residents prior to study enrolment for item optimization. The test is scored based on comparison of residents' answers against the frequency of answers from a panel of experts. Our expert panel was composed of ten adult ICU physicians with more than five years of ICU experience. Questions with consensus and large disagreement among experts were discarded to optimize discriminative power and to minimize measurement error.[18] The final version of the test was delivered in paper format. We used a T-transformation of residents' scores based on the *distribution of panelists' scores* with a mean (standard deviation [SD]) of 80 (5) to facilitate score interpretation.[14] Based on the literature, mean scores of medical students are expected to be around 70 (or 2 effect sizes below expert scores) and mean scores of residents are

expected to reach 75 (or 1 effect size below expert scores).[17]

Part 2: Part-task trainer simulation-based assessment of procedural skills

We trained two groups of four and five raters to use the OSATS-GRS (see ESM, eAppendix 2) to score resident performance during independent completion of two procedures on part-task trainers (simulation-based performance): BMV and ultrasound-guided CVC insertion. The raters included non-procedural experts with an ICU clinical background (e.g., ICU nurse or respiratory therapist). Our goal was to rely on non-physicians to conduct the assessment to facilitate implementation. Raters attended a four-hour group training session during which ten video-recordings of variable performance levels (five per procedure) were reviewed, scored individually, and then discussed as a group until the raters agreed on scale interpretation.

Part 3: Simulated family member encounter-based assessment of communication skills

We designed three versions of a clinical scenario for which two simulated family meetings needed to be conducted: the first meeting was held immediately after a simulated patient admission to the ICU (part A: initial update post-ICU admission) and the second meeting was held later in the ICU course to address goals of care for that same patient (part B: end-of-life discussion). A similar scenario had previously been piloted with multiple groups of ICU residents who confirmed its realism and acceptability.[22] We then trained a group of eight standardized patients to role-play the family members for the three versions of the scenario. Training consisted of a half-day of repeated practice of each scenario, with feedback from a senior professional with expertise in training. The simulated family members were also trained as raters (i.e., to use the communication GRS tool—see ESM eAppendix 3), so that they could assess resident communication performance during the simulated family meetings. The four-hour rater training followed the same process as described above for procedural skills rater training.

Study procedures

We collected data during the first four weeks of residents' ICU rotation. On week 1, they completed a demographic questionnaire, and then participated in a two-hour three-part assessment. This assessment included the SCT station (up to 60 min); the procedural station, including a ten-minute BMV assessment and a 20-min CVC insertion assessment; and the communication station, with two 15-min interactions (part A and part B) with the same simulated family member. On week 4, residents completed a similar two-hour multi-station assessment (same version of the SCT and same procedural station, but a different two-part communication scenario also based on a clinical update and end-of-life discussion, but for a different simulated patient). The assessments were conducted on-site by a research nurse in three separate quiet rooms. On a given day, multiple assessments were often conducted in parallel, so the order of the stations for each resident was randomly assigned. At the end of the last assessment on week 4, residents completed a satisfaction questionnaire regarding the acceptability and fairness of the assessment process. Residents then received written feedback on their performance on the three-station assessment.

To account for resident schedule variability (e.g., post-call days, mandatory off-site weekly specialty-based teaching), we scheduled the competency testing up to one week after the ideal schedule as needed (e.g., on week 2 for pre-rotation testing and on week 5 for post-rotation testing when possible).

Study outcomes

To assess the feasibility and acceptability of the multi-competency assessment, we collected data on the proportion of eligible residents who were recruited and who completed their follow-up, reasons for refusal to participate and for incomplete follow-up, assessment costs, and participants' perceptions of the acceptability and fairness of the competency testing (see ESM, eAppendix 5). To assess residents' learning outcomes, we collected participants' SCT scores for cognitive reasoning, OSATS scores for the two procedural skills, and GRS scores for part A and part B of the communication scenarios at week 1 and week 4.

Statistical analyses

We summarized descriptive data (completion rate, acceptability, assessment scores) and expressed continuous variables as mean (SD) or median [interquartile range] as appropriate. Categorical variables are summarized as frequencies and percentages. The SCT score was calculated and transformed as described above. The procedural OSATS and communication GRS scores were calculated by averaging the individual item scores of each scale (OSATS: seven items rated on a scale from 1 to 5; communication GRS: five items rated on a scale from 1 to 5).

As this was a pilot study, our sample size was one of convenience. In terms of assessing feasibility, our goal was

to enrol ≥ 70% of the eligible residents in the study, and for > 80% of the enrolled residents to complete at least one standardized assessment. We assessed whether there was evidence of learning in three different domains after a four-week ICU rotation; we tested whether mean within-person changes in continuous scores were different from zero using a paired t-test, and whether there were differences between week 1 and week 4 in percentages scoring ≥ 4 using McNemar's test. A four-fold plot was used to depict the changes in these percentages.[23] The threshold of statistical significance was $P < 0.05$.

Results

Feasibility and acceptability

Between January and October 2017, 63 of 91 eligible residents (69%) were enrolled in the study, of which 58 (92%) completed at least one assessment: 18 on week 1 only, eight on week 4 only, and 32 on both weeks 1 and 4. Reasons for declining consent or for not participating after providing consent included: too nervous (n = 6), too busy clinically (n = 2), sick (n = 1), and recent specialty program assessment (n = 1). Consenting residents who could not be scheduled for one or both assessments (n = 5) were either on post-call days/vacations or had competing teaching activities. Participant characteristics of residents who completed the demographic questionnaire (n = 49) are presented in Table 1.

The entire cost for the 90 assessments conducted at the three sites during the study was CAD 33,800, including CAD 23,900 for initial purchase of part-task trainers, CAD 7,500 for simulated patients/raters at an hourly rate of CAD 44, and CAD 2,400 for CVC kits and other supplies.

The majority of the satisfaction survey respondents (36 participants) agreed or strongly agreed that the three-part assessment was fair (n = 24, 67%), of acceptable duration (n = 25, 70%), and measured important clinical abilities (n = 26, 72%). Approximately one-third of the trainees (n = 13, 36%) agreed or strongly agreed that the assessment guided their self-study and/or motivated them to seek involvement in certain clinical activities during their ICU rotation. Nevertheless, a third (n = 12, 33%) of the participants believed that it was tiring, and a smaller proportion of participants believed that the assessment unacceptably interfered with other clinical activities (n = 6, 17%) or that it was a waste of their time (n = 4, 11%).

Table 1 Participant characteristics

Characteristics	(n=49)
ICU site	
Site 1	23 (47)
Site 2	17 (35)
Site 3	9 (18)
Age (yr)	
≤ 25	1 (2)
26–30	36 (73)
31–35	11 (23)
> 35	1 (2)
Training program	
Anesthesia	11 (23)
Internal medicine	29 (59)
Surgery	9 (18)
Postgraduate year	
1	4 (8)
2	26 (53)
3	15 (31)
4	4 (8)
Previous ICU rotations (months)	
0	22 (45)
1	12 (25)
2	10 (20)
3	4 (8)
> 3	1 (2)

ICU = intensive care unit.

Changes in assessment scores between week 1 and week 4 of ICU rotation

For residents (n = 32) who completed assessments both at weeks 1 and 4, the mean (SD) SCT (cognitive reasoning) scores increased from 61.2 (9.1) in week 1 to 64.3 (11.2) in week 4 (mean difference, 3.1; 95% confidence interval [CI], 2.7 to 3.5; P = 0.01). The mean (SD) OSATS (procedural skills) scores for CVC insertion improved significantly from 3.9 (0.9) in week 1 to 4.4 (0.6) in week 4 (mean difference, 0.6; 95% CI, 0.2 to 0.9; P = 0.001). The mean (SD) scores for BMV also improved from 4.1 (0.7) in week 1 to 4.5 (0.5) in week 4 (mean difference, 0.4; 95% CI, 0.2 to 0.6; $P > 0.001$). The effect size for these changes was between 0.3 and 0.6. Nevertheless, the communication scores for scenario part A and B did not change significantly between week 1 and week 4 [part A: 4.1 (0.7) *vs* 3.9 (0.6); mean difference, −0.3; 95% CI, −0.6 to 0.1; P = 0.11; part B: 4.4 (0.6) *vs* 4.1 (0.7); mean difference, −0.2; 95% CI, −0.6 to 0.2; P = 0.32] (Table 2).

The proportion of residents who obtained an OSATS score ≥ 4 (indicating a high level of performance)

Table 2 Assessment scores at week 1 and week 4 of ICU rotation for residents who completed both assessments (n=32)

Types of assessment	Week 1 mean (SD)	Week 4 mean (SD)	Change in scores	95% CI	P value	Effect size
Script concordance test	61.2 (9.1)	64.3 (11.2)	3.1 (6.5)	2.7 to 3.5	0.013	0.34
OSATS (central venous catheter insertion)	3.9 (0.9)	4.4 (0.6)	0.6 (0.8)	0.2 to 0.9	0.0014	0.59
OSATS (bag-mask ventilation)	4.1 (0.7)	4.5 (0.5)	0.4 (0.5)	0.2 to 0.6	0.0005	0.56
Communication GRS (part A)	4.1 (0.7)	3.9 (0.6)	−0.3 (0.8)	−0.6 to 0.1	0.11	−0.37
Communication GRS (part B)	4.4 (0.6)	4.1 (0.7)	−0.2 (0.9)	−0.6 to 0.2	0.32	−0.36

CI = confidence interval; GRS = global rating scale; ICU = intensive care unit; OATS = objective structured assessment of technical skills; SD = standard deviation

increased significantly for both CVC insertion (P = 0.003) and BMV (P = 0.03) between week 1 and week 4. In contrast, the proportion of residents who obtained communication scores $\geq$ 4 for scenario part A and B did not change significantly during the four-week rotation (P = 0.77 and P = 0.29, respectively). Nevertheless, not all residents completed each assessment at both times. Results of residents who completed the same assessment twice (in week 1 and in week 4) showed that the majority of residents who started their ICU rotation with OSATS (procedural) scores $<$ 4, had scores $\geq$ 4 at the end of the rotation (Figure). For the communication GRS scores, many residents already had scores $\geq$ 4 in week 1, which were either maintained or dropped $<$ 4 in week 4 (Figure).

Discussion

Specialty programs rely predominantly on off-service critical care rotations for their residents to learn how to care for critically ill patients. Nevertheless, the ICU environment poses specific educational challenges related to the nature of ICU clinical activities, work organization, and patient safety concerns. Specialty programs most often ignore which learning and assessment opportunities their residents are likely to encounter during an ICU rotation.[24] This project shows that integrating a standardized assessment of resident competencies is feasible and acceptable in the ICU context—i.e., 63 (69%) of 91 eligible residents were enrolled, with 58 (92%) completing at least one assessment, which is near or above our 70% and 80% feasibility targets, respectively. Our study also suggests that residents improve their clinical reasoning and procedural skills during a four-week ICU rotation.

We designed a multi-station, multi-competency assessment that was heavily informed by existing literature on the assessment of cognitive reasoning, procedural skills, and communication.[14–16,18,19,25,26] The assessment specifically targeted domains of competence and competencies relevant to the ICU setting. We successfully implemented the assessment in three ICUs characterized by different teaching schedules, physical and human resources, and work organizations. In addition, the assessment could be completed without the direct involvement of busy critical care physicians. Nevertheless, only two-thirds (63%) of the eligible residents completed at least one assessment over a four-week period, and only one-third (35%) did the assessment twice. Enrolment was possibly limited by residents' common reluctance to be assessed and by a general anxiety related to the beginning of a new ICU rotation at time of recruitment (on day 1 or 2 of the rotation). Assessment scheduling was complicated by residents' working schedule, which was frequently affected by post-call days, vacations, and specialty-based teaching. Some residents also struggled getting timely coverage of their clinical activities by other members of the ICU team so that they could complete their assessment. Implementation costs were also significant and included the purchase of part-task trainers and remuneration of standardized patients. Assessing residents' competencies once a month (e.g., on week 4 for end-of-rotation competencies or on week 1 for self-learning guidance) instead of twice of month would increase feasibility. Assessment costs could possibly be further minimized by using available resources, such as salary-based or volunteered ICU personnel, recovered expired central line venous catheters, and supplies not usable for clinical activities.

Residents who participated in the study largely perceived the assessment as fair and acceptable. Nevertheless, the time commitment (up to four hours) required to complete the two assessments was perceived as problematic (e.g., a waste of time or interfering with clinical activities) by a small proportion (11% and 17%, respectively) of the trainees. The lack of feedback after the first assessment—planned to measure learning from clinical activities between the two assessments—may

have compromised the perceived educational value of the assessment on week 1.

Based on these findings, we believe that the implementation and wider adoption of this multi-station, multi-competency assessment is feasible in many ICUs to either further study the learning outcomes resulting from an off-service ICU rotation or provide formative feedback to specialty residents rotating through the ICU (as part of an overall competency-based assessment strategy), with some caveats. Asking residents to complete the assessment only once (e.g., near the end of the ICU rotation), providing timely feedback, and ensuring proper clinical cross-coverage are strategies that should be considered to increase completion rate and resident acceptability.

In terms of learning outcomes, the cognitive reasoning and procedural scores increased significantly between the beginning and the end of a four-week ICU rotation. This finding suggests that specialty residents further developed their critical care-related cognitive reasoning, in addition to their CVC insertion and BMV skills during a short ICU rotation. Nevertheless, the absolute changes in SCT and OSATS scores are small, and the effect sizes are considered small to moderate for an educational intervention.[27] Based on the literature, SCT effect sizes between residents and experts have been in the range of 0–3.2.[17] Differences in transformed mean scores between residents and experts are typically around 5. We therefore believe that the changes observed after a four-week ICU rotation resulting in effect sizes between 0.34 and 0.59 and in an absolute difference of 3 in transformed SCT scores reflect a meaningful, although modest educational impact. Changes in the categorical distribution of the OSATS scores between week 1 and 4 also support a qualitative difference in the procedural performance of the participants. Given that the trainees did not receive feedback after the assessment conducted on week 1, this change may have resulted from clinical exposure and practice, formal teaching, and/or self-study. Interestingly, the communication scores did not improve during a four-week ICU rotation. Multiple reasons may explain this result. First, junior trainees may not learn communication skills because involvement in complex ICU family meetings is lacking, communication skills are difficult and time-consuming to teach, and these skills take time to develop.[28] Second, residents may become emotionally exhausted, burned out, and lose empathy as the result of their ICU exposure. Third, the communication assessment may not have been responsive to changes in performance, perhaps because we failed to fully engage residents in the second simulated family encounter that was similar to the one conducted in week 1. These results are exploratory and should be investigated further. Better understanding the nature, scope, and inter-individual variability of learning that result from short clinical rotations during residency is key in the context of CBME. The inadequate end-of-rotation competence after one or two months of training among a significant proportion of trainees supports the argument against purely time-based postgraduate training.

Limitations

To determine the acceptability of the assessment, we used a resident survey that was not formally validated prior to this study. In addition, this study may not have been powered to provide narrow estimates of or to detect differences in certain learning outcomes; results should be interpreted with caution. Small sample sizes are a common problem in medical education research.[29] Our intention is to build on this pilot work to conduct a larger study. We concluded that clinical learning was likely the cause of residents' improvement in cognitive reasoning and procedural performance on week 4, but we cannot exclude that residents learned by doing the assessment on week 1.[30] Given the literature on SCT test-retest reliability[31,32] and the importance of feedback for simulation-based learning,[33–36] we believe that this is unlikely. Additional validity evidence based on a larger sample of specialty residents rotating in the ICU is needed before the assessment is used for summative purposes. The assessment cannot be considered a substitute for workplace-based assessments, although both assessment strategies could be complementary. The role of physician *vs* non-physician assessors should also be further examined as results may have differed if ICU physicians had assessed the residents' performance. Lastly, the studied assessment addresses specific ICU-related competencies and results should not be interpreted as a surrogate of specialty residents' "overall ICU competence" because not all competency domains/competencies were assessed.

Conclusions

This study showed that a multi-station, multi-competency assessment for specialty residents rotating in different ICUs is feasible and could be completed by a majority of residents who found the assessment acceptable, especially if costs, time commitment, and interference with clinical activities are minimized. These results, although preliminary, suggest that specialty residents' cognitive reasoning and procedural skills improved, whereas their communication skills were maintained during a four-week ICU rotation. Larger scale studies and additional validity evidence supporting this assessment program will help better define learning and assessment opportunities and

outcomes for specialty residents completing short ICU rotations.

Author contributions *Dominique Piquette*, *Chris Parshuram*, *Christie Lee*, and *Alberto Goffi* contributed to all aspects of this manuscript, including study conception and design; acquisition, analysis and interpretation of data; and drafting the article. *Catharine Walsh*, *Ryan Brydges*, and *Briseida Mema* contributed to the conception and design of the study, and to the interpretation of data.

Acknowledgement The authors would like to thank the Canadian Institutes for Health Research for their generous support for this project through a project grant (bridge funding).

Disclosures None.

Funding statement Canadian Institutes of Health Research (CIHR) - Project grant (bridge funding). All the authors are current recipients of CIHR grant(s). Dominique Piquette, Ryan Brydges, and Catharine Walsh are currently receiving grants from the Royal College of Physicians and Surgeons of Canada. Chris Parshuram has a patent unrelated to this work (Bedside pediatric early warning system—US 8,550,996; 2013).

Editorial responsibility This submission was handled by Dr. Hilary P. Grocott, Editor-in-Chief, *Canadian Journal of Anesthesia.*

References

1. *Hoffman KG*, *Donaldson JF*. Contextual tensions of the clinical environment and their influence on teaching and learning. Med Educ 2004; 38: 448-54.
2. *Minter DJ*, *Levy SD*, *Rao SR*, *Currier PF*. Intensive care unit rotations and predictors of career choice in pulmonary/critical care medicine: a survey of internal medicine residency directors. Crit Care Res Pract 2018; DOI: https://doi.org/10.1155/2018/9496241.
3. *Piquette D*, *Moulton CA*, *LeBlanc VR*. Balancing care and teaching during clinical activities: 2 contexts, 2 strategies. J Crit Care 2015; 30: 678-84.
4. *Piquette D*, *Moulton CA*, *LeBlanc VR*. Creating learning momentum through overt teaching interactions during real acute care episodes. Adv Health Sci Educ Theory Pract 2015; 20: 903-14.
5. *Bannister SL*, *Dolson MS*, *Lingard L*, *Keegan DA*. Not just trust: factors influencing learners' attempts to perform technical skills on real patients. Med Educ 2018; 52: 605-19.
6. *Holmboe E*, *Ginsburg S*, *Bernabeo E*. The rotational approach to medical education: time to confront our assumptions? Med Educ 2011; 45: 69-80.
7. *Piquette D*, *Moulton CA*, *LeBlanc VR*. Model of interactive clinical supervision in acute care environments. Balancing patient care and teaching. Ann Am Thorac Soc 2015; 12: 498-504.
8. *Barsuk JH*, *Cohen ER*, *Feinglass J*, *McGaghie WC*, *Wayne DB*. Residents' procedural experience does not ensure competence: a research synthesis. J Grad Med Educ 2017; 9: 201-8.
9. *Ottestad E*, *Boulet JR*, *Lighthall GK*. Evaluating the management of septic shock using patient simulation. Crit Care Med 2007; 35: 769-75.
10. *Schroedl CJ*, *Corbridge TC*, *Cohen ER*, *et al*. Use of simulation-based education to improve resident learning and patient care in the medical intensive care unit: a randomized trial. J Crit Care 2012; 27(219): e7-13.
11. *Singer BD*, *Corbridge TC*, *Schroedl CJ*, *et al*. First-year residents outperform third-year residents after simulation-based education in critical care medicine. Simul Healthc 2013; 8: 67-71.
12. *Royal College of Physicians and Surgeons of Canada.* Competence by Design. Available from URL: http://www.royalcollege.ca/rcsite/cbd/competence-by-design-cbd-e (accessed August 2020).
13. *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: 560-75.
14. *Charlin B*, *Gagnon R*, *Lubarsky S*, *et al*. Assessment in the context of uncertainty using the script concordance test: more meaning for scores. Teach Learn Med 2010; 22: 180-6.
15. *Hatala R*, *Cook DA*, *Brydges R*, *Hawkins R*. Constructing a validity argument for the objective structured assessment of technical skills (OSATS): a systematic review of validity evidence. Adv Health Sci Educ Theory Pract 2015; 20: 1149-75.
16. *Hodges B*, *McIlroy JH*. Analytic global OSCE ratings are sensitive to level of training. Med Educ 2003; 37: 1012-6.
17. *Dory V*, *Gagnon R*, *Vanpee D*, *Charlin B*. How to construct and implement script concordance tests: insights from a systematic review. Med Educ 2012; 46: 552-63.
18. *Fournier JP*, *Demeester A*, *Charlin B*. Script concordance tests: guidelines for construction. BMC Med Inform Decis Mak 2008; DOI: https://doi.org/10.1186/1472-6947-8-18.
19. *Lubarsky S*, *Charlin B*, *Cook DA*, *Chalk C*, *van der Vleuten CP*. Script concordance testing: a review of published validity evidence. Med Educ 2011; 45: 329-38.
20. *Lubarsky S*, *Dory V*, *Duggan P*, *Gagnon R*, *Charlin B*. Script concordance testing: from theory to practice: AMEE guide no. 75. Med Teach 2013; 35: 184-93.
21. *Lubarsky S*, *Gagnon R*, *Charlin B*. Scoring the script concordance test: not a black and white issue. Med Educ 2013; 47: 1159-61.
22. *Kalocsai C*, *Amaral A*, *Piquette D*, *et al*. "It's better to have three brains working instead of one": a qualitative study of building therapeutic alliance with family members of critically ill patients. BMC Health Serv Res 2018; DOI: https://doi.org/10.1186/s12913-018-3341-1.
23. *Friendly M*. A fourfold display for 2 by 2 by *k* tables. Technical Report 217: Psychology Department, York University; 1994.
24. *Cao S*, *Ladowski S*, *Goffi A*, *et al*. What do non-critical care residents actually learn during an intensive care unit rotation: time to find out! Can J Anesth 2019; 66: 874-6.
25. *LeBlanc VR*, *Tabak D*, *Kneebone R*, *Nestel D*, *MacRae H*, *Moulton CA*. Psychometric properties of an integrated assessment of technical and communication skills. Am J Surg 2009; 197: 96-101.
26. *Moulton CA*, *Tabak D*, *Kneebone R*, *Nestel D*, *MacRae H*, *LeBlanc VR*. Teaching communication skills using the integrated procedural performance instrument (IPPI): a randomized controlled trial. Am J Surg 2009; 197: 113-8.
27. *Hojat M*, *Xu G*. A visitor's guide to effect sizes: statistical significance versus practical (clinical) importance of research findings. Adv Health Sci Educ Theory Pract 2004; 9: 241-9.
28. *Curtis JR*, *Back AL*, *Ford DW*, *et al*. Effect of communication skills training for residents and nurse practitioners on quality of communication with patients with serious illness: a randomized trial. JAMA 2013; 310: 2271-81.
29. *Cook DA*, *Beckman TJ*. Reflections on experimental research in medical education. Adv Health Sci Educ Theory Pract 2010; 15: 455-64.

30. *Larsen DP*, *Butler AC*, *Roediger HL 3rd*. Test-enhanced learning in medical education. Med Educ 2008; 42: 959-66.
31. *Ahmadi SF*, *Khoshkish S*, *Soltani-Arabshahi K*, *et al*. Challenging script concordance test reference standard by evidence: do judgments by emergency medicine consultants agree with likelihood ratios? Int J Emerg Med 2014; DOI: https://doi.org/10.1186/s12245-014-0034-3.
32. *Park AJ*, *Barber MD*, *Bent AE*, *et al*. Assessment of intraoperative judgment during gynecologic surgery using the script concordance test. Am J Obstet Gynecol 2010; 203(240): e1-6.
33. *McGaghie WC*, *Harris IB*. Learning theory foundations of simulation-based mastery learning. Simul Healthc 2018; 13: S15-20.
34. *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: 706-11.
35. *McGaghie WC*, *Issenberg SB*, *Petrusa ER*, *Scalese RJ*. A critical review of simulation-based medical education research: 2003-2009. Med Educ 2010; 44: 50-63.
36. *McGaghie WC*, *Issenberg SB*, *Petrusa ER*, *Scalese RJ*. Revisiting 'a critical review of simulation-based medical education research: 2003-2009'. Med Educ 2016; 50: 986-91.