# Picking the Right Tool for the Job: A Reliability Study of 4 Assessment Tools for Central Venous Catheter Insertion

Jason A. Lord, MSc, MD, FRCPC
Danny J. Zuege, MSc, MD, FRCPC
Maria Palacios Mackay, DDS, MSc, PhD
Amanda Roze des Ordons, MMed, MD, FRCPC
Jocelyn Lockyer, MHA, PhD

## ABSTRACT

**Background** Determining procedural competence requires psychometrically sound assessment tools. A variety of instruments are available to determine procedural performance for central venous catheter (CVC) insertion, but it is not clear which ones should be used in the context of competency-based medical education.

**Objective** We compared several commonly used instruments to determine which should be preferentially used to assess competence in CVC insertion.

**Methods** Junior residents completing their first intensive care unit rotation between July 31, 2006, and March 9, 2007, were video-recorded performing CVC insertion on task trainer mannequins. Between June 1, 2016, and September 30, 2016, 3 experienced raters judged procedural competence on the historical video recordings of resident performance using 4 separate tools, including an itemized checklist, Objective Structured Assessment of Technical Skills (OSATS), a critical error assessment tool, and the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE). Generalizability theory (G-theory) was used to compare the performance characteristics among the tools. A decision study predicted the optimal testing environment using the tools.

**Results** At the time of the original recording, 127 residents rotated through intensive care units at the University of Calgary, Alberta, Canada. Seventy-seven of them (61%) met inclusion criteria, and 55 of those residents (71%) agreed to participate. Results from the generalizability study (G-study) demonstrated that scores from O-SCORE and OSATS were the most dependable. Dependability could be maintained for O-SCORE and OSATS with 2 raters.

**Conclusions** Our results suggest that global rating scales, such as the OSATS or the O-SCORE tools, should be preferentially utilized for assessment of competence in CVC insertion.

## Introduction

Many tools have been developed to assess procedural competence.[1–3] It is unclear which of these tools should be preferentially utilized within the context of competency-based medical education. In the absence of psychometrically sound instruments, it is difficult to justify decision-making related to resident competence, progress, and promotion.

An ideal tool should demonstrate evidence of validity and reliability.[4] Itemized checklists and global rating scales recommended for procedural skills assessments have different strengths and weaknesses.[2,5,6] Checklists focus on technical aspects and specific observable behaviors. Since they are intuitive and provide an objective measure, their scores can be erroneously assumed to be more reliable than those from global assessment scales.[7] Additionally, checklists are limited to assessment of the skill for which they are developed, and scores provided by each specific tool require assessment to ensure there is evidence for validity.[2] Global rating scales, such as the Objective Structured Assessment of Technical Skills (OSATS) or the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE), define performance on a global behavior scale or set of subscales.[3,8] These tools may better discriminate expertise. Rather than comparing performance to an arbitrary standard, they avoid central tendencies by setting standards to identify safe, independent performance.[3,9] Judgments made using these global rating scales are more subjective and potentially suffer from bias due to the influence of past interpretations of performance on current assessments.[10] A recent systematic review[2] failed to clearly identify which tool should be preferentially used for procedural skills assessments. Although interrater reliability scores for checklists compared favorably to global rating scales,

DOI: http://dx.doi.org/10.4300/JGME-D-19-00107.1

*Editor's Note: The online version of this article contains the assessment tools and a study design flowchart.*

other variables such as interitem and interstation reliabilities favored global assessments.[2]

Generalizability theory (G-theory) offers an efficient and practical means to assess performance characteristics of these instruments.[11–14] Unlike classical test theory, G-theory estimates the accuracy of generalizing an observed test score to the trainee's true score under multiple conditions by simultaneously measuring error contributed from the participants, assessment tool items, raters, and interactions among these components. This provides a robust reliability assessment, calculated as a dependability index.[11] Furthermore, G-theory enables investigators to optimize the testing environment by performing a decision study to predict how altering variables pertaining to the assessment process (eg, number of items or raters) affects tool dependability.[11]

Given the common use of central venous catheters (CVCs) in the management of unstable patients, competence in CVC insertion has been identified as a key learning objective for trainees in a variety of specialties.[15–19] The purpose of this study was to compare the dependability of 4 different procedural assessment tools for determining competence in CVC insertion.[13]

**What was known and gap**
A variety of instruments are available to determine procedural performance for central venous catheter (CVC) insertion, but it is not clear which ones are more dependable to assess competence.

**What is new**
A comparison was made, using generalizability theory, of the dependability of 4 procedural assessment tools for determining competence in CVC insertion.

**Limitations**
Study participants came from a single institution, limiting generalizability, and low-fidelity mannequins did not allow for evaluation of all components of CVC insertion.

**Bottom line**
Study results suggest that global rating scales, such as the OSATS and O-SCORE, are more dependable than checklists or critical error tools for assessment of competence in CVC insertion.

## Methods

### Setting and Participants

In 2006, junior residents from a variety of training programs who were completing their first rotation in 1 of 2 intensive care units (ICUs) between July 31, 2006, and March 9, 2007, were invited to participate in a study that involved inserting CVCs while being video recorded. In 2006, residents who had previously completed an ICU rotation, were absent on the final day of the rotation, or were on elective rotations from outside centers were excluded from the video-recording study. Participating residents had gained exposure to CVC insertion during an introductory simulation session prior to the start of their ICU rotation in 2006–2007. Informed consent was obtained for all participants at this time.

### Study Procedures

In 2006–2007, we video-recorded all participants performing right-sided, internal, jugular, and subclavian central line insertions on CVC task trainers (Life/form, Nasco, Fort Atkinson, WI) at the end of their ICU rotation. All procedures were performed using the landmark and Seldinger techniques.[20] Standard triple-lumen CVCs, hospital procedure trays, and barrier precautions were utilized.

### Outcome Assessments

Although this study was originally intended to assess CVC insertion only using a checklist tool, a delay in our data analysis allowed the opportunity to include 4 modified, contemporary assessment tools. These tools were applied between June 1 and September 30, 2016, to judge procedural competence on the historical video recordings of resident performance. Modifications from original tools were designed to limit scoring to procedural steps that were directly observable on the video recordings. Aspects of the procedure that were not captured on the video recordings were excluded from the tools. These tools, created a priori, included an itemized checklist, a critical error tool, the OSATS, and the O-SCORE (provided as online supplemental material).

The itemized checklist, modified from a previously published tool,[6] was designed to assess performance of 9 key steps in CVC insertion. Each step was weighted equally with a maximum possible score of 9. The critical error tool, modified from a previously published procedural error tool,[5] was designed to identify 6 potential critical errors, which were weighted equally. The maximum (worst) score possible was 6. A previously published OSATS tool was modified to assess 5 distinct domains specifically related to the technical components of CVC insertion.[21] The score on this instrument was defined as the total score for all domains with a maximum possible score of 20. Finally, a modified O-SCORE tool was used to assess performance along a competence continuum using a 5-point global rating scale (with a maximum possible score of 5).[3]

Each tool was reviewed by the 3 raters (J.A.L., A.R.d.O., D.J.Z.) prior to the study to assess content, establish clarity, and ensure consistency in scoring. To optimize interrater reliability, all raters scored 5

nonstudy participants using each tool and met as a group to further establish consistent approaches. The raters collectively agreed on the specific behaviors that comprised a "successful" or "unsuccessful" score for each item on each tool.

Three senior, board-certified intensivists independently scored the videos for each resident using each of the 4 assessment tools. All raters were faculty members in an academic critical care medicine department, experienced in assessing resident performance. Raters differed in their prior exposure to the individual assessment tools, and at the time of the study, none of the assessment tools were routinely used clinically by any of the raters. The sequence in which raters used the tools to score the residents was not defined. Raters independently reviewed each video for as long as necessary to complete their assessment. Videos were excluded from the analysis if technical difficulties with the equipment or the video-recording process precluded scoring of the participants.

The original and modified studies were approved by the University of Calgary Conjoint Health Research Ethics Board.

### Data Analysis

Standard deviations were calculated for scores on the assessment tools. Standard error measures were calculated for dependability coefficients. Categorical and continuous descriptive data are presented as proportions, means, and SDs as appropriate. The G-theory was used to estimate the relative contribution of resident performance (our measure of interest for each test) to the test scores, compared with the contribution from measurement error.[11] Our object of interest was the participants; potential sources of measurement error included the raters, examination items, and the interactions between raters-items, participants-raters, participants-items, and participants-items-raters, including unmeasurable error. A 2-facet, fully crossed design was used in which each participant was assessed by each rater on each examination item. Each participant was scored by each rater using all assessment tools, allowing estimation of variance in the observed score contributed by participants, raters, items, and interactions among these variables.

The relative proportion of variance explained by each component was calculated for each assessment tool. Sources of variance evaluated included participants, raters, examination items, and the interactions among them, including unmeasurable error. To provide a standard setting with relevance for competence assessment, each test score was interpreted in an absolute manner. A pass/fail cut score was established for each test using a modified Angoff method.[22] These cut scores were used for decision-making, and the dependability index was calculated for each tool. This coefficient ranges from 0 to 1 and provides a measure of the extent to which score consistency is affected by absolute error.[13] Compared with higher-stakes encounters, such as summative assessments or licensing examinations, we accepted a dependability index threshold of 0.7, which is generally considered adequate for lower-stakes and formative assessments, such as those associated with clinical procedures.[23,24]

The generalizability study (G-study) was followed by a decision study to estimate changes in the dependability of scores as a result of increasing/decreasing the number of raters and/or examination items[11] for the purpose of identifying strategies to improve dependability of assessment tool scores and optimize efficiency.

Data were analyzed using SPSS 24 (IBM Corp, Armonk, NY) for descriptive and inferential statistics.

## Results

### Demographics

One hundred twenty-seven residents rotated through the ICUs during the study period, with 77 (61%) meeting inclusion criteria, and 55 of those 77 residents (71%) agreeing to participate (study design flowchart with enrollment details is provided as online supplemental material). Demographics are provided in TABLE 1.

### Performance Scores

A summary of the resident scores using the 4 assessment tools is provided in TABLE 2. Assessment scores generally centered on the midpoint of the assessment scales, with substantial variation for all tools. The number of videos that were excluded from scoring due to technical difficulties is depicted in the study-design layout (provided as online supplemental material).

### Dependability of Performance Scores

The relative proportion of variance explained by each component was calculated for each assessment tool. Sources of variance evaluated were participants, raters, examination items, and all interactions among them. Summaries of the variance components from each assessment tool for the internal jugular and subclavian sites are provided in TABLE 3. For the checklist and critical error tools, substantial measurement error contributed to the overall variance in the observed test scores. At both CVC sites, measurement

**TABLE 1**
Participant Demographic Information

| Demographic | Total (N = 55), No. % |
|---|---|
| Training program | |
| Family medicine, urban | 20 (36) |
| Internal medicine | 8 (15) |
| Family medicine, rural | 7 (13) |
| Anesthesia | 5 (9) |
| Orthopedic surgery | 4 (7) |
| General surgery | 3 (5) |
| Otolaryngology | 2 (4) |
| Neurology | 2 (4) |
| Emergency medicine | 1 (2) |
| Neurological surgery | 1 (2) |
| Missing | 1 (2) |
| Other | 1 (2) |
| Training year | |
| Postgraduate year 1 | 5 (9) |
| Postgraduate year 2 | 45 (82) |
| Postgraduate year 3 | 4 (7) |
| Missing | 1 (2) |
| Gender | |
| Female | 29 (53) |
| Male | 25 (45) |
| Missing | 1 (2) |

error resulting from the interaction among participant-by-rater-by-item contributed to nearly half of the observed test scores. In contrast, the largest source of variance contributing to observed test scores for the OSATS and O-SCORE at both insertion sites was derived from the participants themselves.

A summary of the dependability coefficients and standard errors of measurement for each assessment tool is in TABLE 4. Scores from the OSATS and O-SCORE tools were more dependable than those from the checklist or critical error tools. Dependability scores were consistent between insertion sites for all tools.

In the decision study, the effect of increasing or decreasing the number of raters or items on each of the assessment tools is summarized in TABLE 5. Our results suggest that adding more examination items to the tools would add relatively little value to improving score reliability. Adding more raters would improve the performance of both the checklist and critical error tools but would not be necessary for the OSATS or O-SCORE tools to meet our dependability-index threshold. Furthermore, the dependability of both the OSATS and O-SCORE could be maintained with one less rater for both insertion sites.

## Discussion

We found that global rating instruments (OSATS and O-SCORE) were more dependable than checklists or critical error tools for assessing procedural competency in CVC, a finding supported by the literature.[2,25,26] Our G-study analysis sheds light on why these tools performed differently.

For the checklist and critical error tools, substantial measurement error contributed to the overall variance in the observed test scores. At both anatomical sites, measurement error resulting from the interaction among participant-by-rater-by-item contributed to nearly half of the observed test scores. This variance reflects the inconsistency resulting from the 3-way interaction among participants, raters, and items,

**TABLE 2**
Participant Scores on Central Venous Catheter Insertion Using All Assessment Tools

| Assessment Tool | n | Maximum Possible Score | Participant Minimum Score | Participant Maximum Score | Participant Mean Score | Participant SD |
|---|---|---|---|---|---|---|
| Checklist | | | | | | |
| Internal jugular | 50 | 9 | 2.7 | 9 | 7 | 1.6 |
| Subclavian | 54 | 9 | 1.5 | 9 | 7.4 | 1.7 |
| Critical error tool | | | | | | |
| Internal jugular | 53 | 6 | 0 | 4.7 | 1.2 | 1.2 |
| Subclavian | 54 | 6 | 0 | 4.7 | 0.9 | 1 |
| OSATS | | | | | | |
| Internal jugular | 53 | 20 | 5 | 20 | 12.2 | 4.2 |
| Subclavian | 53 | 20 | 5.7 | 20 | 15.4 | 3.4 |
| O-SCORE | | | | | | |
| Internal jugular | 53 | 5 | 1 | 5 | 3.1 | 1.2 |
| Subclavian | 53 | 5 | 1 | 5 | 3.7 | 1.1 |

Abbreviations: OSATS, Objective Structured Assessment of Technical Skills; O-SCORE, Ottawa Surgical Competency Operating Room Evaluation.

**TABLE 3**
Variance Percentages for All Assessment Tools

| Source of Variation | Checklist, % | Critical Error, % | OSATS, % | O-SCORE, % |
|---|---|---|---|---|
| Internal jugular site | | | | |
| *p* | 11.4 | 13.3 | 58.5 | 65.5 |
| *r* | 1.1 | 1.6 | 1.5 | 3 |
| *i* | 6.3 | 11.7 | 0 | N/A |
| $p \times r$ | 1.7 | 1.6 | 16.2 | N/A |
| $p \times i$ | 26.1 | 21.8 | 4.0 | N/A |
| $r \times i$ | 0.6 | 1.1 | 1.1 | N/A |
| *pri, e* | 52.8 | 48.9 | 18.7 | N/A |
| *pr, e* | | | | 31.5 |
| Total | | | | 100 |
| Subclavian site | | | | |
| *p* | 12.6 | 14.8 | 42.8 | 54.2 |
| *r* | 0.7 | 0.7 | 1.7 | 2.5 |
| *i* | 2.22 | 5.6 | 0 | N/A |
| $p \times r$ | 3.7 | 0 | 26.5 | N/A |
| $p \times i$ | 27.4 | 19.7 | 4.2 | N/A |
| $r \times i$ | 3 | 2.8 | 0.8 | N/A |
| *pri, e* | 50.4 | 56.3 | 24.0 | N/A |
| *pr, e* | | | | 43.3 |
| Total | | | | 100 |

Abbreviations: *p*, participants; *r*, raters; *i*, examination items; *e*, unmeasured sources of error; OSATS, Objective Structured Assessment of Technical Skills; O-SCORE, Ottawa Surgical Competency Operating Room Evaluation; N/A, not applicable.

confounded by unmeasured sources of variation. These unmeasured sources of variation may relate to random events or sources of variability not measured in the study (eg, time of day, rater mood, room lighting, etc).[11] Measurement error from the interaction of the participant-by-item was the second-largest contributor to the observed test scores for these tools.

**TABLE 4**
Dependability Coefficients for All Assessment Tools

| Assessment Tool | Dependability Index | SEM |
|---|---|---|
| Checklist tool | | |
| Internal jugular | 0.64 | 0.11 |
| Subclavian | 0.65 | 0.10 |
| Critical error tool | | |
| Internal jugular | 0.59 | 0.13 |
| Subclavian | 0.66 | 0.10 |
| OSATS tool | | |
| Internal jugular | 0.88 | 0.29 |
| Subclavian | 0.78 | 0.32 |
| O-SCORE tool | | |
| Internal jugular | 0.85 | 0.45 |
| Subclavian | 0.78 | 0.50 |

Abbreviations: SEM, standard error measure; OSATS, Objective Structured Assessment of Technical Skills; O-SCORE, Ottawa Surgical Competency Operating Room Evaluation.

This variance represents the inconsistency in the trainees' scores across the individual items in the scoring tools.[27] These findings suggest that successful performance in one construct of interest does not necessarily equate to success in other constructs, or overall competence,[25] and reinforces that checklist tool scores frequently have lower reliability than those from global rating scales.[2] Variance contributed by the raters was negligible, suggesting our rater-training process was effective. In contrast, the largest source of variance contributing to the observed test scores for the OSATS and O-SCORE tools was derived from the participants themselves.

Our decision study further supports the use of OSATS and O-SCORE in the assessment of CVC insertion competence. In contrast to the checklist or critical error tool, the dependability coefficients for the OSATS and O-SCORE exceeded our predefined threshold, allowing us to predict ways to improve efficiency without sacrificing dependability. Based on this study, we would be able to maintain the dependability of scores for both anatomical sites using either the OSATS or O-SCORE with one less rater.

Study findings are limited due to our use of a small number of learners from 2 ICUs in a single health care system. Although CVC techniques tend to be

**TABLE 5**
Decision Study Results for All Assessment Tools

| Examination | Dependability Index | −2 Raters | −1 Rater | +1 Rater | +1 Item | +2 Items | +3 Items |
|---|---|---|---|---|---|---|---|
| Checklist tool, IJ | 0.64 | 0.48 | 0.59 | 0.66 | 0.66 | 0.67 | 0.69 |
| Checklist tool, SC | 0.65 | 0.48 | 0.60 | 0.68 | 0.67 | 0.68 | 0.70 |
| Critical error tool, IJ | 0.59 | 0.44 | 0.54 | 0.62 | 0.62 | 0.64 | 0.67 |
| Critical error tool, SC | 0.66 | 0.50 | 0.61 | 0.68 | 0.69 | 0.72 | 0.74 |
| OSATS tool, IJ | 0.88 | 0.72 | 0.83 | 0.90 | 0.88 | 0.89 | 0.89 |
| OSATS tool, SC | 0.78 | 0.56 | 0.71 | 0.82 | 0.79 | 0.79 | 0.80 |
| O-SCORE tool, IJ | 0.85 | 0.65 | 0.79 | 0.88 | N/A | N/A | N/A |
| O-SCORE tool, SC | 0.78 | 0.54 | 0.70 | 0.83 | N/A | N/A | N/A |

Abbreviations: IJ, internal jugular; SC, subclavian; OSATS, Objective Structured Assessment of Technical Skills; O-SCORE, Ottawa Surgical Competency Operating Room Evaluation; N/A, not applicable.

relatively uniform worldwide,[28–31] generalizations to other institutions or different learner groups may be limited. The low-fidelity mannequins did not allow us to evaluate some components of CVC insertion, such as patient positioning, anesthetic instillation, or radiographic confirmation of line position. Technical difficulties with the mannequins prevented some residents from successfully completing the procedure, although the number of studies excluded for analysis on that basis was very small (2%–9%). Since our study assessed resident performance on CVC insertion using simulated task-trainer mannequins, extrapolation to performance on actual patients may be limited. We did not examine resident performance using ultrasound guidance for CVC insertion; therefore, the findings may not be generalized to procedures in which ultrasound is routinely utilized. Although ultrasound is recommended for CVC insertion, this is not always an option.[32] Availability of ultrasound is restricted to clinical sites and departments that can afford the technology, not infrequently limiting access in rural, smaller, or less economically advantaged centers. Furthermore, ultrasound may not be readily available during emergency situations, such as those encountered on some wards or outpatient settings. For these reasons, clinical practitioners often must be able to perform CVC insertion with and without ultrasound guidance. Finally, although scores provided by the OSATS and O-SCORE were more dependable than those from the checklist or critical error tools, 2 raters were still needed to exceed our dependability threshold. This limits their utility as workplace-based assessment instruments. The raters in this study were intensivists in a single academic department, limiting generalizability to raters with different backgrounds and experiences in resident assessment.

Future studies will examine the performance and dependability of the O-SCORE for other clinical procedures, such as ultrasound-guided CVC insertion. This will enable us to determine how broadly this tool can be utilized for procedural skills assessment in critical care medicine.

## Conclusion

Our results suggest that global rating scales, such as the OSATS or O-SCORE, are more dependable than checklists or critical error tools for assessment of competence in CVC insertion, an essential procedural skill for many clinical trainees to acquire. These results, together with the simplicity of the O-SCORE, support adoption of the O-SCORE for assessing CVC insertion at our institution.

## References


1. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective Structured Assessment of Technical Skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273–278. doi:10.1046/j.1365-2168.1997.02502.x.
2. Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ*. 2015;49(2):161–173. doi:10.1111/medu.12621.
3. 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*. 2012;87(10):1401–1407. doi:10.1097/ACM.0b013e3182677805.
4. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach*. 2011;33(3):206–214. doi:10.3109/0142159X.2011.551559.
5. Ma IW, Pugh D, Mema B, Brindle ME, Cooke L, Stromer JN. Use of an error-focused checklist to identify

incompetence in lumbar puncture performances. *Med Educ*. 2015;49(10):1004–1015. doi:10.1111/medu.12809.
6. Ma IW, Zalunardo N, Pachev G, Beran T, Brown M, Hatala R, et al. Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Adv Health Sci Educ Theory Pract*. 2012;17(4):457–470. doi:10.1007/s10459-011-9322-3.
7. Cohen DS, Colliver JA, Robbs RS, Swartz MH. A large-scale study of the reliabilities of checklist scores and ratings of interpersonal and communication skills evaluated on a standardized-patient examination. *Adv Health Sci Educ Theory Pract*. 1996;1(3):209–213. doi:10.1007/BF00162917.
8. 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(5):1149–1175. doi:10.1007/s10459-015-9593-1.
9. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306(9):978–988. doi:10.1001/jama.2011.1234.
10. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17(3):163–170. doi:10.1179/jmt.2009.17.3.163.
11. Shavelson RJ, Webb NM. *Generalizability Theory: A Primer*. Newbury Park, CA: Sage Publications; 1991:137.
12. Lawson DM. Applying generalizability theory to high-stakes objective structured clinical examinations in a naturalistic environment. *J Manipulative Physiol Ther*. 2006;29(6):463–467. doi:10.1016/j.jmpt.2006.06.009.
13. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012;34(11):960–992. doi:10.3109/0142159X.2012.703791.
14. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. *Med Educ*. 2010;44(1):109–117. doi:10.1111/j.1365-2923.2009.03425.x.
15. Accreditation Council for Graduate Medical Education. ACGME program requirements for graduate medical education in critical care medicine. 2011. https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/142_critical_care_medicine_2017-07-01.pdf. Accessed May 28, 2019.
16. Royal College of Physicians and Surgeons of Canada. Objectives of training in the specialty of general surgery. 2017. http://www.royalcollege.ca/rcsite/documents/ibd/general-surgery-otr-e. Accessed May 28, 2019.
17. Royal College of Physicians and Surgeons of Canada. Objectives of training in the specialty of internal medicine. 2011. http://www.royalcollege.ca/rcsite/documents/ibd/internal_medicine_otr_e.pdf. Accessed May 28, 2019.
18. Royal College of Physicians and Surgeons of Canada. Objectives of training in the specialty of emergency medicine. 2014. http://www.royalcollege.ca/rcsite/documents/ibd/emergency_otr_e.pdf. Accessed May 28, 2019.
19. Royal College of Physicians and Surgeons of Canada. Objectives of training in the subspecialty of adult critical care medicine. 2014. http://www.royalcollege.ca/rcsite/documents/ibd/critical_care_adult_otr_e.pdf. Accessed May 28, 2019.
20. Bannon MP, Heller SF, Rivera M. Anatomic considerations for central venous cannulation. *Risk Manag Healthc Policy*. 2011;4:27–39. doi:10.2147/RMHP.S10383.
21. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg*. 1997;173(3):226–230. doi:10.1016/S0002-9610(97)89597-9.
22. Norcini JJ. Setting standards on educational tests. *Med Educ*. 2003;37(5):464–469.
23. Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof*. 2003;23(1):4–12. doi:10.1002/chp.1340230103.
24. Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ*. 2003;326(7388):546–548. doi:10.1136/bmj.326.7388.546.
25. Wimmers PF, Fung CC. The impact of case specificity and generalisable skills on clinical performance: a correlated traits-correlated methods approach. *Med Educ*. 2008;42(6):580–588. doi:10.1111/j.1365-2923.2008.03089.x.
26. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med*. 1999;74(10):1129–1134.
27. Creswell JW. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. 4th ed. Thousand Oaks, CA: SAGE Publications Inc; 2014.
28. Bishop L, Dougherty L, Bodenham A, Mansi J, Crowe P, Kibbler C, et al. Guidelines on the insertion and management of central venous access devices in adults. *Int J Lab Hematol*. 2007;29(4):261–278. doi:10.1111/j.1751-553X.2007.00931.x.
29. Bodenham A. AAGBI safe vascular access guidelines—a reply. *Anaesthesia*. 2016;71(12):1498–1499. doi:10.1111/anae.13719.
30. American Society of Anesthesiologists Task Force on Central Venous Access, Rupp SM, Apfelbaum JL, Blitt C, Caplan RA, Connis RT, et al. Practice guidelines for

central venous access: a report by the American Society of Anesthesiologists Task Force on Central Venous Access. *Anesthesiology*. 2012;116(3):539–573. doi:10.1097/ALN.0b013e31823c9569.

31. Frykholm P, Pikwer A, Hammarskjold F, Larsson AT, Lindgren S, Lindwall R, et al. Clinical guidelines on central venous catheterisation: Swedish Society of Anaesthesiology and Intensive Care Medicine. *Acta Anaesthesiol Scand*. 2014;58(5):508–524. doi:10.1111/aas.12295.
32. Leung J, Duffy M, Finckh A. Real-time ultrasonographically-guided internal jugular vein catheterization in the emergency department increases success rates and reduces complications: a randomized, prospective study. *Ann Emerg Med*. 2006;48(5):540–547. doi:10.1016/j.annemergmed.2006.01.011.

All authors are with Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada. **Jason A. Lord, MSc, MD, FRCPC,** is Clinical Associate Professor of Medicine, Departments of Critical Care Medicine and Emergency Medicine; **Danny J. Zuege, MSc, MD, FRCPC,** is Clinical Professor of Medicine, Department of Critical Care Medicine; **Maria Palacios Mackay, DDS, MSc, PhD,** is Dean of Faculty of Health Science, Department of Family Medicine, San Sebastián University, Chile; **Amanda Roze des Ordons, MMed, MD, FRCPC,** is Clinical Assistant Professor, Departments of Critical Care Medicine and Anesthesia and Division of Palliative Care Medicine; and **Jocelyn Lockyer, MHA, PhD,** is Professor Emerita, Department of Community Health Sciences.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

This work was presented at the Association for Medical Education in Europe Conference, Helsinki, Finland, August 26–30, 2017; and at the International Conference on Residency Education, Québec, Québec, Canada, October 17–22, 2017.

The authors would like to thank Tyrone Donnon, PhD, Department of Community Health Sciences, Cumming School of Medicine, University of Calgary.

Corresponding author: Jason A. Lord, MSc, MD, FRCPC, University of Calgary, Peter Lougheed Centre, Unit 28, 3500 26 Avenue NE, Calgary, AB T1Y 6J4 Canada, 403.943.2569, lordj@ucalgary.ca

Received February 11, 2019; revision received May 4, 2019; accepted May 8, 2019.