Consensus-Based Expert Development of Critical Items for Direct Observation of Point-of-Care Ultrasound Skills

Irene W. Y. Ma, MD, PhD
Janeve Desy, MD, MEHP, RDMS
Michael Y. Woo, MD
Andrew W. Kirkpatrick, MD, MHSc
Vicki E. Noble, MD, FACEP

ABSTRACT

Background Point-of-care ultrasound (POCUS) is increasingly used in a number of medical specialties. To support competency-based POCUS education, workplace-based assessments are essential.

Objective We developed a consensus-based assessment tool for POCUS skills and determined which items are critical for competence. We then performed standards setting to set cut scores for the tool.

Methods Using a modified Delphi technique, 25 experts voted on 32 items over 3 rounds between August and December 2016. Consensus was defined as agreement by at least 80% of the experts. Twelve experts then performed 3 rounds of a standards setting procedure in March 2017 to establish cut scores.

Results Experts reached consensus for 31 items to include in the tool. Experts reached consensus that 16 of those items were critically important. A final cut score for the tool was established at 65.2% (SD 17.0%). Cut scores for critical items are significantly higher than those for noncritical items (76.5% ± SD 12.4% versus 53.1% ± SD 12.2%, $P < .0001$).

Conclusions We reached consensus on a 31-item workplace-based assessment tool for identifying competence in POCUS. Of those items, 16 were considered critically important. Their importance is further supported by higher cut scores compared with noncritical items.

Introduction

Point-of-care ultrasound (POCUS) is increasingly being integrated into patient care in many specialties, such as emergency medicine,[1,2] critical care,[3–5] anesthesiology,[6–8] and internal medicine.[9,10] To support competency-based education,[11] training programs need to establish a programmatic approach to assessments.[12] Recurrent workplace-based observations are essential to help trainees achieve competence and to support decision-making and judgments regarding their competence.[13,14] To date, multiple assessment tools for POCUS skills have been published, with varying amounts of validity evidence to support the interpretation of scores.[15–23] Assessment tools are primarily checklists, global rating scales, or a combination of both. While data suggested that reliability measures and sensitivity to expertise may be higher for global rating scales,[24,25] in the hands of untrained raters, checklists may be easier and more intuitive to use.[26,27] However, checklists risk "rewarding thoroughness," allowing the successful completion of multiple trivial items while masking the commission of a single serious error.[27–31] As such, there is a need to establish which checklist items are critical in POCUS, such that incompetent performances are appropriately identified.

This study sought to develop a consensus-based assessment tool for POCUS skills and to determine which items are critical for competence.

DOI: http://dx.doi.org/10.4300/JGME-D-19-00531.1

Methods

Assessment Tool Construction

Draft assessment items were collated by 2 authors (I.W.Y.M. and V.E.N.) based on a review of the relevant literature regarding directly observed POCUS assessments.[16,19,32–40] Items were then grouped according to key domains (introduction/patient interactions, use of the ultrasound machine, choice of scans, image acquisition, image interpretation, and clinical integration). For each item, respondents were asked its importance for inclusion into a rating tool, and whether learners must successfully complete that item to be considered competent in POCUS (yes, critical; no, noncritical). Importance was rated using a 3-point Likert scale (1, marginal; 2, important; 3, essential to include). This draft survey was then reviewed by all coauthors for item relevance and completeness. It was subsequently piloted for survey content, clarity, and flow on 5 faculty members who taught POCUS in an educational setting (1 emergency physician, 1 general internist, 2 surgeons, and 1 anatomist) and 2 postgraduate year-5 internal medical residents who had completed 1 month of POCUS training. This piloted survey became the instrument used in the first round of the consensus process.

Consensus Process

Between August and December 2016, using a modified Delphi technique,[41] we conducted 3 rounds of an online survey to establish consensus from an expert panel of diverse POCUS specialists and sought their input on the draft assessment items identified in the prior construction stage. Specifically, we sought to achieve consensus on which of the items should be included in a POCUS assessment tool and which items should be considered critical.

The POCUS experts were identified using nonprobability convenience sampling based on international reputation and recruited via an email invitation. Inclusion criteria included completion of at least 1 year of POCUS fellowship training and/or a minimum of 3 years of teaching POCUS.

Consensus to include was defined as 80% or more experts agreeing that an item was essential or important to include in the tool, and consensus to exclude was 80% or more agreeing that the item was marginal. Similarly, consensus for a critical item was defined as 80% or more agreeing that the item must be successfully completed to be considered competent. Items for which the experts had not reached consensus but had $\geq$ 70% agreement were readdressed in subsequent rounds in which items were rated in a binary fashion (yes, should include versus no, should not include).

Standard Setting

To set cut scores for the tool to distinguish between POCUS performances that are competent from performances that are not competent, we invited 12 experts to attend a 3-hour standards setting meeting on March 6, 2017, either in person or via teleconferencing. For this meeting, $\geq$ 50% of these subject matter experts had to have been new (ie, did not participate in the initial expert panel).

At the start of the meeting, we oriented experts to the standards-setting task involved (modified, iterative Angoff method).[42,43] Experts then discussed the behaviors of a borderline POCUS performance to establish a shared mental model of minimally competent performances, defined as those performed unsupervised and considered minimally acceptable. For each item, experts anonymously estimated the percentage of minimally competent POCUS learners who would complete the item successfully. In other words, on an item level, experts were asked to consider a group of 100 borderline learners and estimate how many would successfully complete the item. Experts were blinded to whether or not the item was previously determined by the consensus process to be critically important. Modification to the Angoff method included the use of an iterative process: items with large variances (SD $\geq$ 25%) were discussed and readdressed in subsequent rounds.[44] We decided in advance that no more than 3 rounds of standards setting would be conducted. The final cut score for the entire tool was then derived from the mean of individual-item expert estimates. The final cut score for the critical items was derived from the mean critical-item expert estimates.

This study was approved by the University of Calgary Conjoint Health Research Ethics Board.

What was known and gap
Point-of-care ultrasound (POCUS) is increasingly used in a number of medical specialties. Workplace-based assessments are essential, and there is a need to establish what checklist items are critical when assessing POCUS skills.

What is new
A consensus-based assessment tool for POCUS skills was developed.

Limitations
The tool provides guidance on which assessment items are critically important; it does not specify to educators how a learner must successfully complete that item.

Bottom line
Consensus was reached on a 31-item workplace-based assessment tool for identifying competence in POCUS, with 16 items considered critically important.

Statistical Analysis

Standard descriptive statistics were used in this study. Comparisons of measures between groups were performed using Student's t tests. A 2-sided P value of .05 or less was considered to indicate statistical significance. All analyses were conducted using SAS version 9.4 (SAS Institute Inc, Cary, NC).

Results

Of the 27 experts invited to the panel, 25 (93%) agreed to participate. Their baseline characteristics are presented in TABLE 1.

Assessment Tool

All 25 experts (100%) completed round 1. Experts reached consensus for 31 of the 32 items (97%)

TABLE 1
Baseline Characteristics of Expert Panels for Assessment Tool Construction and Standard Setting

Baseline Characteristics	Consensus-Based Tool Construction, n (%)	Standards Setting Process, n (%)
Total number of experts	25 (100)	12 (100)
Specialty[a]		
Cardiology	2 (8)	0 (0)
Emergency medicine	14 (56)	8 (67)
Internal medicine	8 (32)	2 (17)
Critical care medicine	3 (12)	2 (17)
Pediatric emergency medicine	1 (4)	0 (0)
Surgery	0 (0)	1 (8)
Gender		
Male	20 (80)	7 (58)
Female	5 (20)	5 (42)
Location of practice		
United States of America	18 (72)	8 (67)
California	3 (17)	0 (0)
Colorado	1 (6)	1 (8)
Maine	0 (0)	1 (8)
Massachusetts	3 (17)	2 (17)
Minnesota	2 (11)	1 (8)
North Carolina	1 (6)	0 (0)
New York	1 (6)	0 (0)
Ohio	1 (6)	2 (17)
Oregon	2 (11)	1 (8)
Pennsylvania	1 (6)	0 (0)
South Carolina	2 (11)	0 (0)
Texas	1 (6)	0 (0)
Canada	7 (28)	4 (33)
Alberta	0 (0)	1 (8)
British Columbia	2 (29)	0 (0)
New Brunswick	1 (14)	0 (0)
Ontario	4 (57)	3 (25)
Years of point-of-care ultrasound experience, y		
3–4	1 (4)	0 (0)
5–6	3 (12)	2 (17)
7–8	2 (8)	2 (17)
9–10	3 (12)	2 (17)
More than 10	16 (64)	6 (50)
Completed $\geq$ 1 y of ultrasound fellowship training	16 (64)	9 (75)

[a] Participants were allowed to choose more than 1 option.

for inclusion. The remaining item "Ensures machine charged when not in use" was readdressed in round 2.

The experts reached consensus for 14 of the 32 items (44%) in round 1 as being critically important. The group also reached consensus for 2 additional items as not being critical ("Ensures machine charged when not in use" and "Scans with efficiency of hand motion"). Experts did not reach consensus for critical importance on the remaining 16 of 32 items (50%).

Round 2 was completed by 24 of the experts (96%). For the item "Ensures machine charged when not in use," only 10 of the 24 (42%) felt it should be included in the tool. That item was dropped and not considered further.

In round 2, consensus was achieved on the critical importance of 1 of the 15 items (7%) that the group had not reached consensus on in round 1; 20 of the 24 experts (83%) would fail the learner who does not "appropriately clean the machine and transducers." The 2 items that had $\geq$ 70% agreement for being critical ("Able to undertake appropriate next steps in the setting of unexpected or incidental findings" and "Explains procedure—explain ultrasound, its role, and images—where applicable") were readdressed in round 3.

Round 3 was completed by 22 of the 25 experts (88%) who reached consensus on the item "Able to undertake appropriate next steps in the setting of unexpected or incidental findings" as being critically important (18 of 22, 82%). The group did not achieve consensus on the item "Explains procedure—explain ultrasound, its role, and images—where applicable" (16 of 22, 73%).

The final 31 items included into the assessment tool and the 16 determined to be critical are listed in TABLE 2.

Standards Setting

Twelve experts participated in the standards-setting exercise (TABLE 1). Of those, 6 (50%) served in the panel on tool construction.

In round 1, cut scores were established for 27 of the 31 items (87%). Four items with an SD $\geq$ 25% were discussed and readdressed in round 2 ("Washes hands," "Appropriately enters patient identifier," "Appropriately cleans machine and transducers," "Able to ensure safety of transducers"). After discussion and rerating of those 4 items in round 2, only 1 item continued to have an SD $\geq$ 25% ("Able to ensure safety of transducers"). In round 3 postdiscussion, that item achieved an SD < 25% (mean 42.8% $\pm$ SD 24.1%).

Final cut score of the tool was established at 65.2% $\pm$ SD 17.0% (TABLE 2). Cut scores for critical items are significantly higher than those for noncritical items (76.5% $\pm$ SD 12.4% versus 53.1% $\pm$ 12.2%, $P < .0001$). Cut scores for critical items were also significantly higher than the cut score for the full assessment tool ($P = .022$).

Discussion

In this study, using consensus group methods,[45] our experts agreed on 31 items to be included in the workplace-based POCUS assessment tool. POCUS is a complex skill, involving image acquisition, image interpretation, and clinical integration of findings at the bedside.[46] Our tool included items on those domains.[16,46] In addition, it included items emphasizing the importance of appropriate patient interactions as part of POCUS competence,[47] serving to articulate for educators the breadth of key tasks relevant to the assessment of bedside POCUS skills.

Of the 31 items on the tool, only 16 (52%) were felt to be critically important. Although critical items on clinical and procedural skills have previously been published,[30,48–51] to our knowledge, they have not been established for general POCUS skills. Delineating what items are critical is important for POCUS for 2 reasons. First, POCUS is a relatively new skill. For general medicine, its role has only recently been officially recognized.[9] Having few faculty trained in this skill continues to be the most significant barrier to curriculum implementation for general medicine.[52,53] In Canada, only approximately 7% of internal medicine faculty[54] and 30% of family medicine physicians are trained in POCUS.[55] Without trained faculty, appropriate assessment of trainee skills is highly challenging. Critical items can help guide faculty development efforts by helping them better focus on key essential tasks, thereby more effectively managing rater workload[56] and improving rater performance.[57] Secondly, using key items in assessments may potentially result in higher diagnostic accuracy[30,51] and superior reliability measures,[58] training, and patient safety.[29]

In the era of competency-based medical education,[11] mastery-based learning is associated with improved clinical outcomes.[59,60] Achievement of minimum passing scores set by an expert panel is associated with superior skills and patient outcomes.[61–63] While expert panel cut scores are commonly used for standards setting, others have argued that traditional standards-setting methods result in learners being able to miss a fixed percentage of assessment items, without attention to which items were being missed, resulting in patient safety concerns.[29] We have noted similar concerns in procedural skill assessments in which learners may achieve very high checklist scores, despite having committed serious procedural errors.[27,31] In our present study, we first established which items were considered critical by consensus-group methods. We then applied standards-setting procedures to evaluate cut scores. Blinded to whether or not an item was considered critical, our expert panel's established cut scores for critical items were significantly higher than for noncritical items, suggesting those items may indeed be qualitatively different. Specifically, critical items dealt with key skills in image acquisition (items 7, 9, 14, and 16; TABLE 2), interpretation (items 17, 20, 24, 25, and 26), and safe patient management, such as clinical integration (items 27, 28, 30, and 31),

TABLE 2
Final 31-Item Assessment Tool: Critical Items and Established Cut Scores

Item	Critical Items[a]	Expert Estimate %[b] (SD)
Introduction		
1. Introduces self where applicable (ie, if not already known to patient, patient not critically ill)		72.8 (20.4)
2. Explains procedure (explains ultrasound, its role, and images) where applicable (ie, patient not critically ill)		74.2 (16.1)
3. Washes hands		49.0 (17.8)
4. Ensures patient appropriately and discreetly exposed		55.3 (22.1)
5. Explains ultrasound findings appropriately (even if unsure of results), where applicable	Yes (1)	74.6 (18.1)
Appropriate use of the machine		
6. Appropriately positions the machine		54.3 (19.6)
7. Appropriately applies basic knobology (eg, on/off, depth, gain)	Yes (1)	86.7 (14.8)
8. Appropriately uses examination presets		52.5 (24.8)
9. Chooses correct transducer	Yes (1)	90.0 (14.1)
10. Appropriately enters patient identifier		43.2 (15.7)
11. Able to store relevant images and clips	Yes (1)	61.3 (21.5)
12. Appropriately cleans machine and transducers	Yes (2)	42.1 (16.3)
13. Able to ensure safety of transducers (eg, not dropping transducers)		42.8 (24.1)
Choice of scans based on clinical relevance		
14. Conducts the appropriate types of scans	Yes (1)	80.8 (14.0)
15. Conducts scans in the appropriate prioritization/sequence		64.1 (23.2)
16. Applies appropriate clinical reasoning in choice of scans	Yes (1)	70.1 (10.2)
Image acquisition		
17. Attains minimal criteria	Yes (1)	84.2 (16.1)
18. Positions patient appropriately for specific scans		60.1 (18.6)
19. Scans with adequate transducer pressure		56.5 (19.0)
20. Scans adequately through the entire area of interest	Yes (1)	78.8 (19.8)
21. Able to optimize image appropriately when necessary		42.1 (17.6)
22. Adjusts focal zone appropriately (where relevant and available)		32.5 (18.0)
23. Scans with efficiency of hand motion		37.8 (20.6)
Image interpretation		
24. Able to recognize key findings	Yes (1)	88.3 (11.1)
25. Able to recognize when images are inadequate/insufficient for a given indication	Yes (1)	87.1 (20.5)
26. Recognizes relevant artifacts	Yes (1)	68.3 (19.1)
Scan integration/clinical decision making		
27. Able to determine when and what additional imaging studies/investigations are necessary	Yes (1)	82.2 (17.4)
28. Able to appropriately determine patient disposition based on ultrasound findings	Yes (1)	79.2 (16.9)
29. Able to appropriately incorporate test characteristics (eg, sensitivity/specificity/likelihood ratios) into clinical decision making		60.0 (17.5)
30. Able to appropriately manage unexpected or unknown findings on bedside ultrasound	Yes (3)	67.9 (17.5)
31. Overall, able to determine appropriate next clinical steps	Yes (1)	83.3 (12.1)
Final cut score for the 31-item tool		65.2 (17.0)
Final cut score for the 16 critical item tool		76.5 (12.4)

[a] Critical items are those that the experts indicated that learner should fail the competency-based assessment if the item was not perform satisfactorily; the numbers in parentheses indicate the round in which consensus for the critical item was achieved.
[b] Expert estimate % refers to the expert estimated percentage of borderline learners who would successfully complete the item.

communication of findings (items 5 and 11), and infection-control issues (item 12).

Our study has some limitations. While our tool provides guidance on which assessment items are critically important, it does not specify to educators how a learner must successfully complete that item. For example, the item "Attains minimal criteria" still requires that the faculty be able to recognize what images are of sufficient quality such that image interpretation is even possible. Therefore, faculty training will continue to be an important part of trainee assessments. Further, despite knowing which items are critical, at present, there is no clear guidance on how to assess those items. Three options have been proposed. From a patient-safety perspective, many feel that learners should be required to successfully complete *all* critical items to be considered competent.[64] However, while this approach is appealing from a patient-safety perspective, it may result in greater consequences for the learner. Thus, the defensibility of that approach will require additional validity-evidence data to support its use. For example, evidence demonstrating that raters can rate those items with high interrater reliability would be helpful.[65] A second approach involves setting separate cut scores for critical items than for noncritical items (in the same manner as our present study).[64] Finally, a third approach involves applying item weights,[65] which may be challenging because experts may not agree on what weights to apply. Certainly, within our study, despite iterative discussions, the final variance on some items remained wide, suggesting disagreements among experts. Future studies should determine which of those 3 methods is superior in delineating competent performances from incompetent ones.

Conclusions

Our experts agreed on 31 items for inclusion in a workplace-based assessment tool for POCUS. Of those, 16 (52%) were felt to be critical in nature, with significantly higher cut scores than those for noncritical items. For determining competency in directly observed POCUS skills, faculty should pay particular attention to those items and ensure that they are completed successfully.

References

1. American College of Emergency Physicians. Ultrasound guidelines: emergency, point-of-care and clinical ultrasound guidelines in medicine. *Ann Emerg Med*. 2017;69(5):e27–e54. doi:10.1016/j.annemergmed.2016.08.457.
2. Olszynski P, Kim D, Chenkin J, Rang L. The core emergency ultrasound curriculum project: a report from the Curriculum Working Group of the CAEP Emergency Ultrasound Committee. *CJEM*. 2018;20(2):176–182. doi:10.1017/cem.2017.44.
3. Expert Round Table on Ultrasound in ICU; Cholley BP, Mayo PH, Poelaert J, Vieillard-Baron A, Vignon P, et al. International expert statement on training standards for critical care ultrasonography. *Intensive Care Med*. 2011;37(7):1077–1083. doi:10.1007/s00134-011-2246-9.
4. Mayo PH, Beaulieu Y, Doelken P, Feller-Kopman D, Harrod C, Kaplan A, et al. American College of Chest Physicians/La Société de Réanimation de Langue Française statement on competence in critical care ultrasonography. *Chest*. 2009;135(4):1050–1060. doi:10.1378/chest.08-2305.
5. Arntfield RT, Millington SJ, Ainsworth CD, Arora R, Boyd J, Finlayson G, et al. Canadian recommendations for critical care ultrasound training and competency. *Can Respir J*. 2014;21(6):341–345. doi:10.1155/2014/216591.
6. Neal JM, Brull R, Horn JL, Liu SS, McCartney CJ, Perlas A, et al. The second American Society of Regional Anesthesia and Pain Medicine evidence-based medicine assessment of ultrasound-guided regional anesthesia: executive summary. *Reg Anesth Pain Med*. 2016;41(2):181–194. doi:10.1097/AAP.0000000000000331.
7. Sites BD, Chan VW, Neal JM, Weller R, Grau T, Koscielniak-Nielsen ZJ, et al; American Society of Regional Anesthesia and Pain Medicine; European Society of Regional Anaesthesia and Pain Therapy Joint Committee. The American Society of Regional Anesthesia and Pain Medicine and the European Society of Regional Anaesthesia and Pain Therapy Joint Committee recommendations for education and training in ultrasound-guided regional anesthesia. *Reg Anesth Pain Med*. 2009;34(1):40–46. doi:10.1097/AAP.0b013e3181926779.
8. Meineri M, Bryson GL, Arellano R, Skubas N. Core point-of-care ultrasound curriculum: what does every anesthesiologist need to know? *Can J Anesth*. 2018;65(4):417–426. doi:10.1007/s12630-018-1063-9.
9. American College of Physicians. ACP statement in support of point-of-care ultrasound in internal medicine. https://www.acponline.org/meetings-courses/focused-topics/point-of-care-ultrasound/acp-statement-in-support-of-point-of-care-ultrasound-in-internal-medicine. Accessed February 4, 2020.
10. Ma IWY, Arishenkoff S, Wiseman J, Desy J, Ailon J, Martin L, et al. Internal medicine point-of-care ultrasound curriculum: consensus recommendations from the Canadian Internal Medicine Ultrasound (CIMUS) group. *J Gen Intern Med*.

2017;32(9):1052–1057. doi:10.1007/s11606-017-4071-5.

11. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638–645. doi:10.3109/0142159X.2010.501190.
12. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676–682. doi:10.3109/0142159X.2010.500704.
13. Lockyer J, Carraccio C, Chan MK, Hart D, Smee S, Touchie C, et al. Core principles of assessment in competency-based medical education. *Med Teach*. 2017;39(6):609–616. doi:10.1080/0142159X.2017.1315082.
14. Harris P, Bhanji F, Topps M, Ross S, Lieberman S, Frank JR, et al. Evolving concepts of assessment in a competency-based world. *Med Teach*. 2017;39(6):603–608. doi:10.1080/0142159X.2017.1315071.
15. Black H, Sheppard G, Metcalfe B, Stone-McLean J, McCarthy H, Dubrowski A. Expert facilitated development of an objective assessment tool for point-of-care ultrasound performance in undergraduate medical education. *Cureus*. 2016;8(6):e636. doi:10.7759/cureus.636.
16. Todsen T, Tolsgaard MG, Olsen BH, Henriksen BM, Hillingsø JG, Konge L, et al. Reliable and valid assessment of point-of-care ultrasonography. *Ann Surg*. 2015;261(2):309–315. doi:10.1097/SLA.0000000000000552.
17. Ziesmann MT, Park J, Unger BJ, Kirkpatrick AW, Vergis A, Logsetty S, et al. Validation of the quality of ultrasound imaging and competence (QUICk) score as an objective assessment tool for the FAST examination. *J Trauma Acute Care Surg*. 2015;78(5):1008–1013. doi:10.1097/TA.0000000000000639.
18. Dyre L, Nørgaard LN, Tabor A, Madsen ME, Sørensen JL, Ringsted C, et al. Collecting validity evidence for the assessment of mastery learning in simulation-based ultrasound training. *Ultraschall Med*. 2016;37(4):386–392. doi:10.1055/s-0041-107976.
19. Hofer M, Kamper L, Sadlo M, Sievers K, Heussen N. Evaluation of an OSCE assessment tool for abdominal ultrasound courses. *Ultraschall Med*. 2011;32(2):184–190. doi:10.1055/s-0029-1246049.
20. American College of Emergency Physicians. Emergency ultrasound standard reporting guidelines. https://www.acep.org/globalassets/uploads/uploaded-files/acep/clinical-and-practice-management/policy-statements/information-papers/emergency-ultrasound-standard-reporting-guidelines—2018.pdf. Accessed February 4, 2020.
21. Schmidt JN, Kendall J, Smalley C. Competency assessment in senior emergency medicine residents for core ultrasound skills. *West J Emerg Med*. 2015;16(6):923–926. doi:10.5811/westjem.2015.9.28587.
22. Skaarup SH, Laursen CB, Bjerrum AS, Hilberg O. Objective and structured assessment of lung ultrasound competence. A multispecialty Delphi consensus and construct validity study. *Ann Am Thorac Soc*. 2017;14(4):555–560. doi:10.1513/AnnalsATS.201611-894OC.
23. Patrawalla P, Eisen LA, Shiloh A, Shah BJ, Savenkov O, Wise W, et al. Development and validation of an assessment tool for competency in critical care ultrasound. *J Grad Med Educ*. 2015;7(4):567–573. doi:10.4300/JGME-D-14-00613.1.
24. Ilgen JS, Ma IWY, Hatala R, Cook DA. A systematic review of validity evidence for checklists vs global rating scales in simulation-based assessment. *Med Educ*. 2015;49(2):161–173. doi:10.1111/medu.12621.
25. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med*. 1999;74(1):1129–1134. doi:10.1097/00001888-199910000-00017.
26. Lammers RL, Davenport M, Korley F, Griswold-Theodorson S, Fitch MT, Narang AT, et al. Teaching and assessing procedural skills using simulation: metrics and methodology. *Acad Emerg Med*. 2008;15(11):1079–1087. doi:10.1111/j.1553-2712.2008.00233.x.
27. Walzak A, Bacchus M, Schaefer JP, Zarnke K, Glow J, Brass C, et al. Diagnosing technical competence in six bedside procedures: comparing checklists and a global rating scale in the assessment of resident performance. *Acad Med*. 2015;90(8):1100–1108. doi:10.1097/ACM.0000000000000704.
28. Cunnington JPW, Neville AJ, Norman GR. The risks of thoroughness: reliability and validity of global ratings and checklists in an OSCE. *Adv Health Sci Educ Theory Pract*. 1996;1(3):227–233. doi:10.1007/BF00162920.
29. Yudkowsky R, Tumuluru S, Casey P, Herlich N, Ledonne C. A patient safety approach to setting pass/fail standards for basic procedural skills checklists. *Simul Healthc*. 2014;9(5):277–282. doi:10.1097/SIH.0000000000000044.
30. Ma IWY, 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.
31. 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.

32. Atkinson P, Bowra J, Lambert M, Lamprecht H, Noble V, Jarman B. International Federation for Emergency Medicine point-of-care ultrasound curriculum guidelines. *CJEM*. 2015;17(2):161–170. doi:10.1017/cem.2015.8.
33. Sisley AC, Johnson SB, Erickson W, Fortune JB. Use of an objective structured clinical examination (OSCE) for the assessment of physician performance in the ultrasound evaluation of trauma. *J Trauma*. 1999;47(4):627–631. doi:10.1097/00005373-199910000-00004.
34. Woodworth GE, Carney PA, Cohen JM, Kopp SL, Vokach-Brodsky LE, Horn JL, et al. Development and validation of an assessment of regional anesthesia ultrasound interpretation skills. *Reg Anesth Pain Med*. 2015;40(4):306–314. doi:10.1097/AAP.0000000000000236.
35. Ziesmann MT, Park J, Unger B, Kirkpatrick AW, Vergis A, Pham C, et al. Validation of hand motion analysis as an objective assessment tool for the Focused Assessment with Sonography for Trauma examination. *J Trauma Acute Care Surg*. 2015;79(4):631–637. doi:10.1097/TA.0000000000000813.
36. Heinzow HS, Friederichs H, Lenz P, Schmedt A, Becker JC, Hengst K, et al. Teaching ultrasound in a curricular course according to certified EFSUMB standards during undergraduate medical education: a prospective study. *BMC Med Educ*. 2013;13:84. doi:10.1186/1472-6920-13-84.
37. Bentley S, Mudan G, Strother C, Wong N. Are live ultrasound models replaceable? Traditional versus simulated education module for FAST exam. *West J Emerg Med*. 2015;16(6):818–822. doi:10.5811/westjem.2015.9.27276.
38. Lam SH, Bailitz J, Blehar D, Becker BA, Hoffmann B, Liteplo AS, et al. Multi-institution validation of an emergency ultrasound image rating scale—a pilot study. *J Emerg Med*. 2015;49(1)32–39.e1. doi:10.1016/j.jemermed.2015.01.010.
39. Amini R, Adhikari S, Fiorello A. Ultrasound competency assessment in emergency medicine residency programs. *Acad Emerg Med*. 2014;21(7):799–801. doi:10.1111/acem.12408.
40. Lewiss RE, Pearl M, Nomura JT, Baty G, Bengiamin R, Duprey K, et al. CORD-AEUS: consensus document for the emergency ultrasound milestone project. *Acad Emerg Med*. 2013;20(7):740–745. doi:10.1111/acem.12164.
41. Dalkey NC. *The Delphi Method: An Experimental Study of Group Opinion*. Santa Monica, CA: RAND Corp; 1969.
42. Angoff W, ed. *Scales, Norms, and Equivalent Scores*. 2nd ed. Washington, DC: American Council on Education; 1971.
43. Hurtz GM, Auerbach MA. A meta-analysis of the effects of modifications to the Angoff method on cutoff scores and judgment consensus. *Educ Psychol Meas*. 2003;63(4):584–601. doi:10.1177/0013164403251284.
44. Ricker KL. Setting cut-scores: a critical review of the Angoff and modified Angoff methods. *Alberta J Educ Res*. 2006;52(1).
45. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. *Med Teach*. 2017;39(1):14–19. doi:10.1080/0142159X.2017.1245856.
46. Soni NJ, Schnobrich D, Matthews BK, Tierney DM, Jensen TP, Dancel R, et al. Point-of-care ultrasound for hospitalists: a position statement of the society of hospital medicine. *J Hosp Med*. 2019;14:e1–e6. doi:10.12788/jhm.3079.
47. Tolsgaard MG, Todsen T, Sorensen JL, Ringsted C, Lorentzen T, Ottesen B, et al. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. *PloS One*. 2013;8(2):e57687. doi:10.1371/journal.pone.0057687.
48. Werner HC, Vieira RL, Rempell RG, Levy JA. An educational intervention to improve ultrasound competency in ultrasound-guided central venous access. *Pediatr Emerg Care*. 2016;32(1):1–5. doi:10.1097/PEC.0000000000000664.
49. Brown GM, Otremba M, Devine LA, Gray C, Millington SJ, Ma IWY. Defining competencies for ultrasound-guided bedside procedures: consensus opinions from Canadian physicians. *J Ultrasound Med*. 2016;35(1):129–141. doi:10.7863/ultra.15.01063.
50. Barsuk JH, McGaghie WC, Cohen ER, Balachandran JS, Wayne DB. Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit. *J Hosp Med*. 2009;4(7):397–403. doi:10.1002/jhm.468.
51. Yudkowsky R, Park YS, Riddle J, Palladino C, Bordage G. Clinically discriminating checklists versus thoroughness checklists: improving the validity of performance test scores. *Acad Med*. 2014;89(7):1057–1062. doi:10.1097/ACM.0000000000000235.
52. Hall J, Holman H, Bornemann P, Barreto T, Henderson D, Bennett K, et al. Point of care ultrasound in family medicine residency programs: a CERA study. *Fam Med*. 2015;47(9):706–711.
53. Schnobrich DJ, Gladding S, Olson APJ, Duran-Nelson A. Point-of-care ultrasound in internal medicine: a national survey of educational leadership. *J Grad Med Educ*. 2013;5(3):498–502. doi:10.4300/JGME-D-12-00215.1.

54. Ailon J, Nadjafi M, Mourad O, Cavalcanti R. Point-of-care ultrasound as a competency for general internists: a survey of internal medicine training programs in Canada. *Can Med Educ J*. 2016;7(2):e51–e69.
55. Micks T, Braganza D, Peng S, McCarthy P, Sue K, Doran P, et al. Canadian national survey of point-of-care ultrasound training in family medicine residency programs. *Can Fam Physician*. 2018;64(1):e462–e467.
56. Tavares W, Eva KW. Impact of rating demands on rater-based assessments of clinical competence. *Educ Prim Care*. 2014;25(6):308–318. doi:10.1080/14739879.2014.11730760.
57. Tavares W, Ginsburg S, Eva KW. Selecting and simplifying: rater performance and behavior when considering multiple competencies. *Teach Learn Med*. 2016;28(1):41–51. doi:10.1080/10401334.2015.1107489.
58. Daniels VJ, Bordage G, Gierl MJ, Yudkowsky R. Effect of clinically discriminating, evidence-based checklist items on the reliability of scores from an internal medicine residency OSCE. *Adv Health Sci Educ Theory Pract*. 2014;19(4):497–506. doi:10.1007/s10459-013-9482-4.
59. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med*. 2013;88(8):1178–1186. doi:10.1097/ACM.0b013e31829a365d.
60. McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ*. 2014;48(4):375–385. doi:10.1111/medu.12391.
61. Barsuk JH, McGaghie WC, Cohen ER, O'Leary KJ, Wayne DB. Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med*. 2009;37(10):2697–2701.
62. Wayne DB, Barsuk JH, O'Leary KJ, Fudala MJ, McGaghie WC. Mastery learning of thoracentesis skills by internal medicine residents using simulation technology and deliberate practice. *J Hosp Med*. 2008;3(1):48–54. doi:10.1002/jhm.268.
63. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Use of simulation-based education to reduce catheter-related bloodstream infections. *Arch Intern Med*. 2009;169(15):1420–1423. doi:10.1001/archinternmed.2009.215.
64. Yudkowsky R, Park YS, Downing SM, eds. *Assessment in Health Professions Education*. 2nd ed. New York, NY: Routledge; 2020.
65. American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.

Irene W.Y. Ma, MD, PhD, is Associate Professor, Department of Medicine, Simulation Lead, W21C, and Co-Program Director, Internal Medicine Residency Program Procedural Skills and Ultrasonography, University of Calgary, Calgary, Alberta, Canada; **Janeve Desy, MD, MEHP, RDMS,** is Clinical Assistant Professor, Department of Medicine, and Co-Program Director; Internal Medicine Residency Program Procedural Skills and Ultrasonography, University of Calgary; **Michael Y. Woo, MD,** is Associate Professor, and Teaching Chair, Distinguished Teachers Program, Undergraduate Medical Education, University of Ottawa, Ottawa, Ontario, Canada, and Fellowship Program Director, Emergency Medicine Ultrasonography, Department of Emergency Medicine, Ottawa Hospital, Ottawa, Ontario, Canada; **Andrew W. Kirkpatrick, MD, MHSc,** is Professor of Surgery and Critical Care Medicine, University of Calgary and Foothills Medical Centre, Calgary, Alberta, Canada; and **Vicki E. Noble, MD, FACEP,** is Vice-Chair for Academic Affairs and Program Director, Center for Emergency Medicine, University Hospitals, Cleveland Medical Center, and Professor, Case Western Reserve School of Medicine.

Funding: This work was funded by a Medical Council of Canada Research in Clinical Assessment grant. The funding source had no role in the design or conduct of the study, its analyses, interpretation of the data, or decision to submit results.

Conflict of interest: Dr Ma is funded as the John A. Buchanan Chair in General Internal Medicine at the University of Calgary. Dr Kirkpatrick has consulted for the Innovative Trauma Care and Acelity companies.

Acknowledgments: We thank the Medical Council of Canada Research for funding this work, the experts who participated in this study, as well as Sydney Haubrich, Julie Babione, and the W21C for their assistance.

Corresponding author: Irene W. Y. Ma, MD, PhD, University of Calgary, 3330 Hospital Drive NW, Calgary, AB T2N 4N1 Canada, 403.210.7369, fax 403.283.6151, ima@ucalgary.ca

Received July 28, 2019; revision received December 11, 2019; accepted December 31, 2019.