

Development and Validation of an Assessment Tool for Competency in Critical Care Ultrasound

Paru Patrawalla, MD
 Lewis Ari Eisen, MD
 Ariel Shiloh, MD
 Brijen J. Shah, MD
 Oleksandr Savenkov, PhD

Wendy Wise, MD
 Laura Evans, MD
 Paul Mayo, MD
 Demian Szyld, MD, EdM

ABSTRACT

Background Point-of-care ultrasound is an emerging technology in critical care medicine. Despite requirements for critical care medicine fellowship programs to demonstrate knowledge and competency in point-of-care ultrasound, tools to guide competency-based training are lacking.

Objective We describe the development and validity arguments of a competency assessment tool for critical care ultrasound.

Methods A modified Delphi method was used to develop behaviorally anchored checklists for 2 ultrasound applications: "Perform deep venous thrombosis study (DVT)" and "Qualify left ventricular function using parasternal long axis and parasternal short axis views (Echo)." One live rater and 1 video rater evaluated performance of 28 fellows. A second video rater evaluated a subset of 10 fellows. Validity evidence for content, response process, and internal consistency was assessed.

Results An expert panel finalized checklists after 2 rounds of a modified Delphi method. The DVT checklist consisted of 13 items, including 1.00 global rating step (GRS). The Echo checklist consisted of 14 items, and included 1.00 GRS for each of 2 views. Interrater reliability evaluated with a Cohen kappa between the live and video rater was 1.00 for the DVT GRS, 0.44 for the PSLA GRS, and 0.58 for the PSSA GRS. Cronbach α was 0.85 for DVT and 0.92 for Echo.

Conclusions The findings offer preliminary evidence for the validity of competency assessment tools for 2 applications of critical care ultrasound and data on live versus video raters.

Introduction

Portable ultrasound technology is increasingly used in the diagnosis and management of critically ill patients. As an easily accessible and frequently used bedside tool, critical care ultrasound (CCUS) has the potential to improve patient safety; however, in the hands of untrained providers, this technology may be harmful. The current standard of procedure logs with arbitrary thresholds and subjective assessment of skill is not adequate to assess competency.^{1,2} This has led to greater interest in developing tools to assess complex procedures such as bronchoscopy that include technical skills, medical knowledge, and clinical judgment.^{3,4}

With the introduction of the new accreditation system⁵ by the Accreditation Council for Graduate Medical Education,⁵ critical care medicine (CCM) and pulmonary/CCM fellowship programs are required to demonstrate trainee competency in procedures, including ultrasound.⁶⁻⁸ Performing ultrasound safely and efficiently is an important

entrustable professional activity for pulmonary medicine and CCM.⁹ Although the specific components required for CCUS competence have been defined,¹⁰ and a framework for training standards developed,¹¹ there is no certification program to ensure uniform standards.

Training standards include didactics in general CCUS and basic critical care echocardiography, practical training on clinically normal volunteers, bedside scanning on patients with a range of conditions, and pathology.¹¹ There is no consensus on the number of scans needed to achieve competence.¹¹ Developing a competency assessment instrument using psychometric principles with strong evidence for validity is essential in the era of outcomes-based education. A structured approach to validity for instrument development has been used for procedural skills.¹²⁻¹⁵ However, we are not aware of any valid and reliable instruments to measure comprehensive CCUS skill.

The purpose of this study was to develop a competency assessment tool for CCUS and provide evidence for its validity.

DOI: <http://dx.doi.org/10.4300/JGME-D-14-00613.1>

Methods

We conducted a prospective study to develop and provide validity evidence for a CCUS competency assessment tool.

The New York University Institutional Review Board approved this study.

Methods for Validity Analysis

We adopted the framework of Downing¹⁶ to provide a validity argument for this tool. Elements include content, response process, internal structure, relationship to other variables, and consequences.

Checklist Development: The instrument was developed by the modified Delphi method.^{17,18} The Delphi method is a consensus-based content generation method frequently used to develop guidelines, policy statements, and assessment tools. The modified Delphi method differs in that an initial structure is given to participants, from which feedback is generated and incorporated until consensus is reached. This method has been used to develop performance checklists with high content-related validity evidence.^{17,18} One author (P.P.), a critical care physician and ultrasound educator, created comprehensive step-by-step procedure checklists for CCUS modules based on existing literature and published guidelines.¹⁰ A modified Delphi method was used to edit and finalize the checklists. In round 1, an expert panel (N = 4) reviewed checklists, made free text edits, and classified step importance using a 9-point Likert scale (1 to 3, not important; 4 to 6, somewhat important; 7 to 9, very important). Steps with a mean rating ≤ 3 were eliminated, and suggested edits were incorporated. Repeated iterations of this method were continued until consensus was reached.

The expert panel included faculty with expertise in CCUS, checklist development, and educational outcome measures. All raters were faculty with extensive experience in CCUS application and education and Delphi panel members. Information regarding the Delphi process is presented descriptively. Means (ranges) are reported for checklist development.

Response Process: We standardized rater activities through the use of scripted instructions. Participants imaged the same healthy actor. Feedback was formalized using a structured report card.

Setting and Participants: First-year CCM fellows from the New York metropolitan area attended a

What was known and gap

Critical care fellows are expected to use point-of-care ultrasound, yet assessment tools to guide development of these skills are lacking.

What is new

Checklists to assess ultrasound examination skills for deep venous thrombosis and left ventricular function as well as comparisons of live and video ratings of performance.

Limitations

Raters were members of an expert Delphi group, limiting generalizability; study did not assess outcomes.

Bottom line

Checklists with acceptable interrater reliability to provide feedback to guide learners' skill development.

regional, 3-day, introductory CCUS course that meets multisociety-recommended training standards¹¹ and includes didactic presentations, image interpretation, and image acquisition on healthy actors. The training included essential CCUS modules as defined by a consensus statement:¹⁰ basic echocardiography, vascular access, vascular diagnostic, pleural, lung, and abdominal ultrasound.

Six to 30 months after attending the course, fellows from 3 training programs completed a comprehensive CCUS assessment. Cognitive skill was assessed with an image-based multiple-choice test. One live and 1 video rater reviewed 28 fellows, and a second video rater reviewed a subset of 10 using the skills checklists. The live rater completed checklists on specific tasks (ie, "Perform DVT study" [DVT] and "Qualify left ventricular function using parasternal long axis and parasternal short axis views" [Echo]). The live rater assigned each task using a script, directly observed the fellow, and then provided checklist-guided feedback. Time was allotted for deliberate practice until competence was achieved. Video raters completed checklists remotely by examining a video recording (1 camera on the probe, 1 on the console, and direct capture of ultrasound images via Arcadia 5.08, Education Management Solutions). Fellow performance on each task was categorized as competent, competent with areas for improvement, or not competent based on the global rating step (GRS). Fellows and program directors received standardized report cards.

Statistics: Cronbach α coefficients were calculated using individual item scores completed by the live rater and 1 video rater on 28 subjects. Only items viewable by video raters were used to assess reliability. Total percent agreement and a Cohen

TABLE 1
Delphi Method for the DVT Checklist

Step	Behavior-Based Anchors	Mean (Range)
1	Uses vascular probe	8 (3–9)
2	Has dot to left of screen	7 (4–9)
3	Places probe in cross-section so that left of probe = left on screen	8 (6–9)
4	Adjusts depth to center femoral vessels on screen	8 (7–9)
5	Adjusts gain so that vascular structures are anechoic and appropriate texture is seen in adjacent muscle	8 (7–9)
6	Identifies CFV	9 (8–9)
7	Identifies greater saphenous vein	9 (8–9)
8	Identifies superficial femoral vein	9 (8–9)
9	Identifies deep femoral vein	8 (6–9)
10	Identifies popliteal vein	9 (8–9)
11	Compresses at all junction points	9 (7–9)
12	Attempts to compress along entire SFV (can be stopped by examiner)	7 (4–8)
13 ^a	<i>Acceptable examination of 1 leg for DVT</i>	...

Abbreviations: DVT, deep venous thrombosis; CFV, common femoral vein; SFV, superficial femoral vein.

^a Steps in italics were added during the first Delphi round and were not rated using a Likert scale.

kappa were used to measure interrater reliability between live and video raters (N = 28) and between 2 video raters (N = 10). The relationship between individual steps and the GRS is presented descriptively. Mean SD is presented for the multiple-choice test. Analysis was performed using SPSS version 22 (IBM Corp).

Relationship to Other Variables: This element requires correlation to an external performance measure. Lack of other existing performance measures for this construct prevented our ability to assess this element.

Consequences: Evidence of consequences was derived from the psychometric data.

Results

Content

Round 1 results of the modified Delphi method are presented for DVT (TABLE 1) and Echo (TABLE 2). All steps had a mean rating > 3 and therefore none were eliminated. Addition of a GRS to each checklist was the only edit. Consensus of 100% was reached with round 2. The final DVT checklist consists of 13

TABLE 2
Delphi Method for the Echo Checklist

Step	Behavior-Based Anchors	Mean (Range)
1	Places patient supine or in left lateral decubitus position	9 (7–9)
2	Uses low-frequency probe in “cardiac” preset	8 (6–9)
3	Dot to right of screen	9 (8–9)
4	Begins with probe marker facing patient’s right shoulder	8 (3–9)
5	Adjust depth so that cardiac chambers are seen in center of screen	9 (9)
6	Adjusts gain so that cardiac chambers are anechoic	9 (8–9)
7	Identifies an appropriate PSLA window (bisects AV and MV in same plane with heart horizontally across screen)	9 (9)
8	Identifies LV, RV, valves, descending aorta ^a	9 (7–9)
9	Rotates probe marker toward patient’s left shoulder for PSSA without “moving” probe to a different position on chest	8 (7–9)
10	Identifies an appropriate PSSA midventricular view (perpendicular slice through LV, not oblique; papillary muscles seen clearly)	8 (7–9)
11	Identifies LV, RV, and papillary muscles ^a	9 (7–9)
12	Correctly identifies LV function (hyperdynamic, normal, mild-moderate dysfunction, or severe dysfunction)	9 (9)
13 ^b	<i>Acceptable parasternal long axis view</i>	...
14 ^b	<i>Acceptable parasternal short axis view</i>	...

Abbreviations: PSLA, parasternal long axis; AV, aortic valve; MV, mitral valve; LV, left ventricle; RV, right ventricle; PSSA, parasternal short axis.

^a Can be prompted to identify structures out loud.

^b Steps in italics were added during the first Delphi round and were not rated using a Likert scale.

dichotomous (yes-no) steps, including the GRS, and the Echo checklist consists of 14 steps, including the GRS for each of 2 Echo views.

Internal Structure

The DVT checklist had a Cronbach α of 0.85 across both live and video raters. The Echo checklist had a Cronbach α of 0.92 for both raters.

Interrater reliability was assessed using percent agreement and a Cohen kappa (TABLES 3 and 4). Steps poorly visualized by the video rater were excluded. Prevalence for some steps was at the extremes (eg, all learners completed the item correctly or no learner

TABLE 3
Interrater Reliability for Each DVT Checklist Step

Step	Live Versus Video Rater (N = 28)		Video Versus Video Rater (N = 10)	
	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)
1. Uses vascular probe	100	^a	100	^a
2. Has dot to left of screen	100	^a	100	^a
6. Identifies CFV	68	0.21 (–0.04–0.46)	50	0.14 (–0.13–0.41)
7. Identifies greater saphenous vein	93	0.79 (0.52–1.00)	70	0.21 (–0.44–0.86)
8. Identifies superficial femoral vein	86	0.62 (0.29–0.95)	80	0.41 (–0.18–1.00)
9. Identifies deep femoral vein	96	0.78 (0.37–1.00)	90	0.62 (–0.05–1.00)
11. Compresses at all junction points	93	0.47 (–0.14–1.00)	100	^a
12. Attempts to compress along entire SFV	82	0.64 (0.37–0.91)	50	0.0 (–0.61–0.61)
13. Acceptable examination of 1 leg for DVT	100	1.00 (1.00)	100	^a

Abbreviations: DVT, deep venous thrombosis; CFV, common femoral vein; SFV, superficial femoral vein.

^a Kappa could not be calculated due to a lack of variation in the data.

completed the item correctly). Kappas for DVT steps ranged from 0.21 to 1.00 between live and video raters and from 0 to 0.62 between 2 video raters (TABLE 3). Total percent agreement between live and video raters and between 2 video raters was 100% for the DVT GRS (TABLE 5). Kappas for Echo steps ranged from 0.29 to 0.58 between live and video raters and from 0.74 to 1.00 between 2 video raters (TABLE 4). Kappas were 0.44 for the GRS of parasternal long axis and 0.58 for parasternal short axis between live

and video raters. Between 2 video raters, kappa was 0.74 for parasternal short axis (TABLE 5).

Performance on DVT steps 6 to 9 was the primary determinant of the DVT GRS. These steps required raters to simultaneously see and hear learner identification of anatomy, which was inconsistently audible on video. Echo step 7 primarily influenced parasternal long axis GRS, and step 10, the parasternal short axis GRS. These steps necessitated careful evaluation of ultrasound images. Video raters were able to replay

TABLE 4
Interrater Reliability for Each Echo Checklist Step

Step	Live Versus Video Rater (N = 28)		Video Versus Video Rater (N = 10)	
	% Agreement	Kappa (95% CI)	% Agreement	Kappa (95% CI)
1. Places patient supine or in left lateral decubitus position	93	0.47 (–0.14–1.00)	90	^a
2. Uses low-frequency probe in “cardiac” preset	100	^a	100	^a
3. Dot to right of screen	93	0.46 (–0.17–1.00)	100	1.00 (1.00)
4. Begins with probe marker facing patient’s right shoulder	100	^a	100	^a
7. Identifies an appropriate PSLA window (<i>bisects AV and MV in same plane with heart horizontally across screen</i>)	79	0.29 (–0.1–0.68)	90	^a
10. Identifies an appropriate PSSA midventricular view (<i>perpendicular slice through LV, not oblique; papillary muscles seen clearly</i>)	82	0.58 (0.27–0.89)	90	0.74 (0.27–1.00)
13. Acceptable parasternal long axis view	82	0.44 (0.05–0.83)	90	^a
14. Acceptable parasternal short axis view	82	0.58 (0.27–0.89)	90	0.74 (0.27–1.00)

Abbreviations: PSLA, parasternal long axis; AV, aortic valve; MV, mitral valve; PSSA, parasternal short axis; LV, left ventricle.

^a Kappa could not be calculated due to a lack of variation in the data.

TABLE 5
Interrater Reliability of Global Rating Steps for DVT and Echo

	Live Versus Video Rater (N = 28)		Video Versus Video Rater (N = 10)	
	Total % Agreement	Kappa (95% CI)	Total % Agreement	Kappa (95% CI)
DVT study	100	1.00 (1.00)	100	^a
Echo (PSLA)	82	0.44 (0.05–0.83)	90	^a
Echo (PSSA)	82	0.58 (0.27–0.89)	90	0.74 (0.27–1.00)

Abbreviations: DVT, deep venous thrombosis; PSLA, parasternal long axis; PSSA, parasternal short axis.

^a Kappa could not be calculated due to a lack of variation in the data.

without restriction, likely accounting for the higher reliability between video raters.

The mean SD score on the cognitive portion of the assessment was 81.9% (10.7), reflecting the skill homogeneity of these learners.

Discussion

We describe a novel tool for assessing CCUS competence with preliminary evidence of its validity using criteria recommended by Downing.¹⁶ Lack of CCUS expertise is an area for improvement in many fellowship programs,¹⁹ and remote asynchronous video review offers an alternative to the time-intensive method of direct supervision. Importantly, our study provides valuable information on the benefits and limitations of live versus video review in CCUS.

Evidence for Validity

We present validity evidence of this instrument following current theories.^{16,20} We offer evidence for content validity, as the authors are content experts in CCUS and educational outcomes assessment. We also standardized rater activities through scripted instructions, offering evidence for response process validity. All participants imaged the same healthy actor, and feedback was formalized through a structured report card. We also offer evidence for internal consistency of both checklists. The Echo GRSs showed moderate interrater reliability, and the DVT GRS showed excellent interrater reliability. With expert raters, GRS may in fact be more reliable than checklists.^{21–23} However, there was variability in the reliability of individual steps. We could not assess relationship to other variables, as the task included no alternative performance measures, nor can we offer evidence of consequences validity. The effect of the assessment on the examinee and learning is often used to determine if the instrument should be used for high-stakes or low-stakes assessment. However, no

high-stakes CCUS examination exists, limiting our ability to assess this element.

Our tool was evaluated in a safe environment on a healthy actor. Fellows were excused from clinical responsibilities and they commented that this provided a unique learning opportunity. The assessment, including a multiple-choice examination, checklists, and deliberate practice, was typically completed within 1 hour. Faculty volunteered their time, and the only cost was for the actor.

Our study had several limitations. Raters were members of the modified Delphi panel, limiting generalizability to raters who are not content experts or did not develop this tool. Tool dissemination will require a rater guide and validation among raters not involved in tool development. Further detail on correct and incorrect task performance should be added to the behavioral anchors to improve interrater reliability.

The extreme prevalence of some items complicates reliability analysis. This relates to the skill homogeneity of subjects (ie, reflected in the similarity of their cognitive examination scores) and the lack of variability in the test subject. Interrater reliability was poor or showed no correlation for several individual checklist items. Several checklist items were excluded as they were poorly seen on video. Video raters also commented that camera footage was inadequate to assess some checklist items.

Future Directions

Implementation of a CCUS competency-based assessment is needed to ensure a new generation of CCUS-competent clinicians. Poor reliability of specific checklist items demonstrates the challenges of remote, asynchronous assessment by video raters. A live rater may best use this tool, particularly given the advantage of providing supervised performance improvement. This tool should be combined with other sources of evaluation for summative assessment or certification.²⁴ This tool needs additional validity evidence among live raters not involved in checklist

development and will also need further evaluation in patients with pathology. Further study should be performed in a heterogeneous group of learners to assess this tool's ability to discriminate learner levels.

Conclusion

We designed a formative assessment tool to establish existing skills, provide checklist-directed feedback, and guide deliberate practice. We also present preliminary validity evidence for a competency assessment tool in critical care ultrasound. This instrument may be used in combination with other measures for formative assessment in critical care medicine fellows.

References

1. Kohls-Gatzoulis JA, Regehr G, Hutchison C. Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. *Can J Surg*. 2004;47(4):277–283.
2. Lund ME. Twenty-five questions: an important step on a critical journey. *Chest*. 2009;135(2):256–258.
3. Davoudi M, Osann K, Colt HG. Validation of two instruments to assess technical bronchoscopic skill using virtual reality simulation. *Respiration*. 2008;76(1):92–101.
4. Quadrelli S, Davoudi M, Galíndez F, Colt HG. Reliability of a 25-item low-stakes multiple-choice assessment of bronchoscopic knowledge. *Chest*. 2009;135(2):315–321.
5. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056.
6. Accreditation Council for Graduate Medical Education. The Internal Medicine Subspecialty Milestones Project. <http://www.acgme.org/acgmeweb/portals/0/pdfs/milestones/internalmedicinesubspecialtymilestones.pdf>. Accessed May 13, 2015.
7. Accreditation Council for Graduate Medical Education. ACGME program requirements for graduate medical education in critical care medicine. http://www.acgme.org/acgmeweb/portals/0/pfassets/2013-pr-faq-pif/142_critical_care_int_med_07132013.pdf. Accessed May 13, 2015.
8. Accreditation Council for Graduate Medical Education. ACGME program requirements for graduate medical education in pulmonary disease and critical care medicine (internal medicine). http://www.acgme.org/acgmeweb/portals/0/pfassets/2013-pr-faq-pif/156_pulmonary_critical_care_int_med_07132013.pdf. Accessed May 13, 2015.
9. Fessler HE, Addrizzo-Harris D, Beck JM, Buckley JD, Pastores SM, Piquette CA, et al. Entrustable professional activities and curricular milestones for fellowship training in pulmonary and critical care medicine: report of a multisociety working group. *Chest*. 2014;146(3):813–834.
10. Mayo PH, Beaulieu Y, Doelken P, Feller-Kopman D, Harrod C, Caplan 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.
11. Expert Round Table on Ultrasound in ICU. International expert statement on training standards for critical care ultrasonography. *Intensive Care Med*. 2011;37(7):1077–1083.
12. Davoudi M, Colt HG, Osann KE, Lamb CR, Mullon JJ. Endobronchial ultrasound skills and tasks assessment tool: assessing the validity evidence for a test of endobronchial ultrasound-guided transbronchial needle aspiration operator skill. *Am J Respir Crit Care Med*. 2012;186(8):773–779.
13. Davoudi M, Osann K, Colt HG. Validation of two instruments to assess technical bronchoscopic skill using virtual reality simulation. *Respiration*. 2008;76(1):92–101.
14. Nielsen DG, Gotzsche O, Eika B. Objective structured assessment of technical competence in transthoracic echocardiography: a validity study in a standardised setting. *BMC Med Educ*. 2013;13:47.
15. Salamonsen M, McGrath D, Steiler G, Ware R, Colt H, Fielding D. A new instrument to assess physician skill at thoracic ultrasound, including pleural effusion markup. *Chest*. 2013;144(3):930–934.
16. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–837.
17. Morgan PJ, Lam-McCulloch J, Herold-McIlroy J, Tarshis J. Simulation performance checklist generation using the Delphi technique. *Can J Anaesth*. 2007;54(12):992–997.
18. Cheung JJ, Chen EW, Darani R, McCartney CJ, Dubrowski A, Awad IT. The creation of an objective assessment tool for ultrasound-guided regional anesthesia using the Delphi method. *Reg Anesth Pain Med*. 2012;37(3):329–333.
19. Eisen LA, Leung S, Gallagher AE, Kvetan V. Barriers to ultrasound training in critical care medicine fellowships: a survey of program directors. *Crit Care Med*. 2010;38(10):1978–1983.
20. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7–e16.
21. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an

- OSCE-format examination. *Acad Med.* 1998;73(9):993–997.
22. 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.
 23. 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.
 24. Schuwirth LW, van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478–485.



At the time of writing, **Paru Patrawalla, MD**, was Assistant Program Director, Fellowship Program in Pulmonary and Critical Care Medicine, and Assistant Professor of Medicine, Department of Medicine/Division of Pulmonary, Critical Care, and Sleep Medicine, New York University School of Medicine, and is now Director, Fellowship Program in Pulmonary and Critical Care Medicine, Department of Medicine/Division of Pulmonary, Critical Care, and Sleep Medicine, Mount Sinai Beth Israel, and Assistant Professor of Medicine, Icahn School of Medicine at Mount Sinai; **Lewis Ari Eisen, MD**, is Director of Education for Critical Care and Associate Professor of Clinical Medicine, Department of Medicine, Division of Critical Care Medicine, Albert Einstein College of Medicine, Montefiore Medical Center; **Ariel Shiloh, MD**, is Director, Critical Care Consult Service, Associate Program Director,

Fellowship Program in Critical Care Medicine, and Assistant Professor of Clinical Medicine and Neurology, Department of Medicine, Division of Critical Care Medicine, Albert Einstein College of Medicine; **Brijen J. Shah, MD**, is Director of Graduate Medical Education for Faculty Development, Quality and Patient Safety, and Assistant Professor of Medicine, Department of Medicine, Henry D. Janowitz Division of Gastroenterology and Brookdale Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai; **Oleksandr Savenkov, PhD**, is Associate Research Scientist, Division of Education Quality and Analytics, Institute for Innovations in Medical Education, New York University School of Medicine; **Wendy Wise, MD**, is Clinical Instructor of Medicine, Department of Medicine, Division of Pulmonary, Critical Care, and Sleep Medicine, New York University School of Medicine; **Laura Evans, MD**, is Associate Program Director, Fellowship Program in Pulmonary and Critical Care, and Associate Professor of Medicine, Department of Medicine, Division of Pulmonary, Critical Care, and Sleep Medicine, New York University School of Medicine; **Paul Mayo, MD**, is Professor of Medicine, Department of Medicine, Division of Pulmonary and Critical Care Medicine, Hofstra Northshore-LIJ School of Medicine; and **Demian Szyld, MD, EdM**, is Medical Director, New York Simulation Center for the Health Sciences, and Assistant Professor of Emergency Medicine, New York University School of Medicine.

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

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

Corresponding author: Paru Patrawalla, MD, Mount Sinai Beth Israel, Division of Pulmonary, Critical Care, and Sleep Medicine, 7 Dazian, First Avenue and 16th Street, New York, NY 10003, 212.420.2892, ppatrawalla@chpnet.org

Received November 3, 2014; revisions received February 14, 2015, and April 5, 2015; accepted May 5, 2015.