Collecting Validity Evidence for Simulation-Based Assessment of Point-of-Care Ultrasound Skills

Jesper Kørup Jensen, MD, Liv Dyre, MB, Mattis Enggaard Jørgensen, MD, Lisbeth Anita Andreasen, MD, Martin Grønnebæk Tolsgaard, MD, PhD

Objectives—The aim of this study was to examine the validity of a simulator test designed to evaluate focused assessment with sonography for trauma (FAST) skills.

Methods—Participants included a group of ultrasound novices (n = 25) and ultrasound experts (n = 10). All participants had their FAST skills assessed using a virtual reality ultrasound simulator. Procedural performance on the 4 FAST windows was assessed by automated simulator metrics, which received a passing or failing score. The validity evidence for these simulator metrics was examined by a stepwise approach according to the Standards for Educational and Psychological Testing. Metrics with validity evidence were included in a simulator test, and the reliability of test scores was determined. Finally, a pass/fail level for procedural performance was established.

Results—Of the initial 55 metrics, 34 (61.8%) had validity evidence ($P < .01$). A simulator test was constructed based on the 34 metrics with established validity evidence, and test scores were calculated as percentages of the maximum score. The median simulator test scores were 14.7% (range, 0%–47.1%) and 94.1% (range, 94.1%–100%) for novices and experts, respectively ($P < .001$). The pass/fail level was determined to be 79.7%.

Conclusions—The performance of FAST examinations can be assessed in a simulated setting using defensible performance standards, which have both good reliability and validity.

Key Words—focused assessment with sonography for trauma; mastery learning; point-of-care ultrasound; simulation-based medical education; validity evidence; virtual reality

Received December 2, 2016, from the Department of Anesthesiology and Intensive Care, Odense University Hospital, Odense, Denmark (J.K.J.); Center of Fetal Medicine, Department of Obstetrics, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark (L.D.); Department of Radiology, Næstved, Slagelse, Ringsted (NSR) Hospitals, Denmark (M.E.J.); and Copenhagen Academy for Medical Education and Simulation, Rigshospitalet, Copenhagen, Denmark (L.A.A., L.D., M.G.T.). Dr Jørgensen is currently with the Department of Radiology, Bispebjerg Hospital, Copenhagen, Denmark. Manuscript accepted for publication March 7, 2017.

We thank Kirsten Engel, MD, who helped with language review of the final draft of the paper. This study was funded by the University of Copenhagen. The funder had no influence on the study design, conduct, or drafting of the manuscript.

Address correspondence to Jesper Kørup Jensen, MD, Department of Anesthesiology and Intensive Care, Odense University Hospital, J.B. Winsløws Vej 4, 5000 Odense C, Denmark.

E-mail: jesperkoerup@gmail.com

Abbreviations
FAST, focused assessment with sonography for trauma

doi:10.1002/jum.14292

Point-of-care ultrasound imaging has increasingly become an integral skill in the surgical specialties.[1] In particular, the focused assessment with sonography for trauma (FAST) examination is widely used during resuscitation of patients with blunt abdominal injuries.[2,3] The aim of the FAST examination is to detect free fluid, which is suggestive of intra-abdominal bleeding, and it represents an invaluable shortcut for the rapid diagnosis of life-threatening intra-abdominal hemorrhage or hemopericardium.[3,4]

Although the sensitivity and specificity of ultrasound findings are highly dependent on the skills of the operator,[5] previous studies have reported short learning curves for the FAST examination.[6–8] However, the exact number of supervised scans needed to achieve acceptable levels of technical performance and diagnostic accuracy is controversial,[6–8] and existing estimates vary between 10 and 200 scans. The wide range reflected in these numbers may indicate that the volume of scans alone is not a good predictor of competence.

Some operators may not be qualified for independent practice after completing the required number of scans, whereas others may become competent long before they are permitted to practice independently. According to best practice in medical education, operators should be judged against a reliable and valid performance criterion to ensure that the individual operator has attained a minimum performance standard before transitioning to independent practice.[9–11] However, training and assessment in a clinical setting place considerable demands on time from both trainees and faculty. Moreover, the availability of cases needed to demonstrate multiple different clinical presentations often limits the speed with which new operators can be trained and certified in the performance of FAST examinations.

The use of virtual reality simulators has gained widespread popularity in multiple specialties and for several types of clinical procedures. Virtual reality ultrasound simulators present novice operators with multiple types of disorders and provide hands-on experience with little involvement from faculty[12–17] because of the use of automated performance assessments (ie, simulator metrics). These simulator metrics can be used for formative feedback as well as to determine when operators have achieved the skills needed for clinical practice. However, built-in simulator metrics are not always valid markers of competence.[17–21] In other words, built-in metrics often fail to discriminate between groups with different levels of competence.[22] The validity and reliability of simulation-based assessments must therefore be examined to rely on their use for evaluating when trainees have acquired certain performance standards. These performance standards must also be established to determine when operators are qualified for clinical practice. Performance standards should be developed and justified so that a simulator test reliably discriminates between those trainees who are fit for subsequent practice with patients and those who are not. The objective of this study was, therefore, to determine the validity and reliability of simulation-based assessments of FAST skills as well as to establish defensible performance standards on a virtual reality ultrasound simulator.

Materials and Methods

Setting and Design

Data collection took place between September 1, 2015, and December 30, 2015, at the Department of Radiology, Slagelse Hospital (Slagelse, Denmark), Næstved Hospital (Næstved, Denmark), and the Copenhagen Academy for Medical Education and Simulation. The Committees on Health Research Ethics of the Capital Region of Denmark deemed the study exempt from ethical review (protocol 15010447). This study involved exploring the validity of simulation-based FAST assessments for novice ultrasound operators. Validity evidence was assessed from different steps according to the Standards for Educational and Psychological Testing, including content evidence, response process, internal structure, relationship with other variables, and test consequences.[22] In accordance with contemporary terminology,[22] we use the term "validity evidence" to stress that we are evaluating the evidence that supports the interpretation of test scores. An overview and short explanations of the different steps in the validation process are provided in Figure 1.

Participant Recruitment

The expert group consisted of consultant radiologists from 2 different university hospitals (Næstved Hospital and Slagelse Hospital) with no previous simulation experience. The novice group included final-year medical students from the University of Copenhagen. None of the novices had any experience with FAST or simulation.

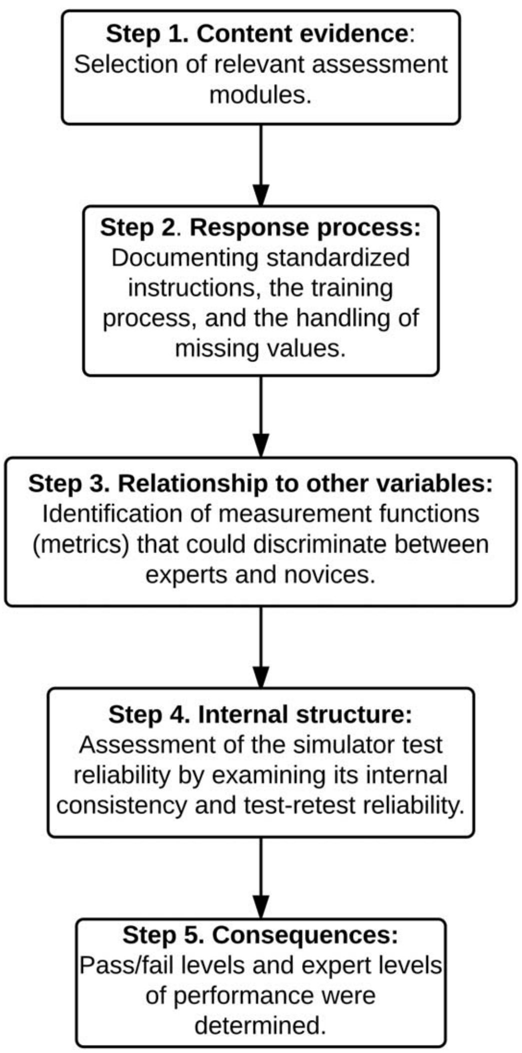
Equipment

All assessment and training was conducted with a transabdominal Scantrainer (Medaphor, Cardiff, Wales; Figure 2). The simulator consists of a haptic device, which provides force feedback, and a personal computer with touch screen functions (Pentium Core i5, Vaio SVL2412M1EB; Sony Corporation, Tokyo, Japan), which displays a 2-dimensional–mode output and basic functions of a standard ultrasound device, such as freeze, capture, caliper, focus, and gain. The haptic device consists of a curvilinear transducer supported by 3 robotic arms, which provide force feedback during scans and allow unlimited movement across a virtual abdomen. An additional screen was placed directly in front of the haptic device. The image on the additional screen showed the anatomic structures being scanned, a 3-dimensional image of the virtual patient, and the placement of the transducer in relation to the patient. The training program consisted of 5 different training modules, each containing a number of tasks (eg, outline the fluid in the Morison pouch). The modules and tasks are displayed in Appendix 1. The tasks were divided into 3 different categories: (1) locate a specific area (eg, locate the left hemidiaphragm and subphrenic space); (2) sweep a specific area (eg, sweep the subphrenic space); and (3) outline the fluid when present (eg, outline the subphrenic fluid) using the caliper function.

Step 1: Content Evidence

In the first step of the validation process, content evidence was examined to ensure that the content of the simulator test reflected what it was supposed to measure. A simulator instructor (J.K.J.) and a medical education researcher (L.D.) identified modules that were of relevance to the FAST examination on the virtual reality simulator (Medaphor Scantrainer). Five training modules specifically designed for training the FAST examination with a total of 55 metrics were selected and served as a single simulator test (Appendix 1). Apart from the presence/absence of free intra-abdominal fluid, there were no differences between the modules with free fluid and the modules without. For this study, only the modules with free fluid were selected.

Figure 1. Flowchart of the validation process.



Step 2: Response Process

The response process, namely, the way in which the participants interacted with the simulator test, was evaluated, and efforts were made to ensure a standardized approach to the assessment process. All participants were instructed to complete the simulator test, which included the 5 training modules. All participants were provided instructions regarding the simulated environment and equipment. All instructions were standardized according to a prespecified protocol. The instructions included a brief presentation of the functions required to complete the tasks. Additionally, the novices were shown basic ultrasound imaging techniques, including orientation of the transducer, manipulation of the transducer (sliding, angling, and rotating the transducer), and specific functions on the simulator (freeze, capture, and caliper). At the end of the introduction, a complete FAST scan on the simulator was demonstrated. With the simulation software provided by the manufacturer, the automated simulator metrics were applied to assess participants' performance while attempting the assignment. Thus, tasks that were not completed according to simulator instructions resulted in missing metric values. A simulator instructor (J.K.J.) was present during the entire test for both groups. A maximum of 40 minutes was allowed to complete the test, and the instructor provided no instructions on how to complete the specific task at hand. The simulator instructor (J.K.J.) provided technical assistance for the novices as needed.

Step 3: Relationship With Other Variables

All simulator metrics were assessed for their ability to discriminate between different levels of operator experience, which in test theory is also known as "relationship with other variables." Simulator metrics were considered to have validity evidence if significant differences were found between the scores of the two groups. Only metrics that had validity evidence were included in the final simulator test.

Step 4: Internal Structure

The distribution and patterns of test scores were evaluated through their internal structure. The internal structure of the final simulator test was evaluated by determining its internal consistency and the test-retest reliability. Internal consistency was assessed for the simulator metrics included in the final test by the Cronbach α statistic. The test-retest reliability was assessed by calculating an intraclass correlation coefficient for the novices' first 2 attempts at the simulator test.

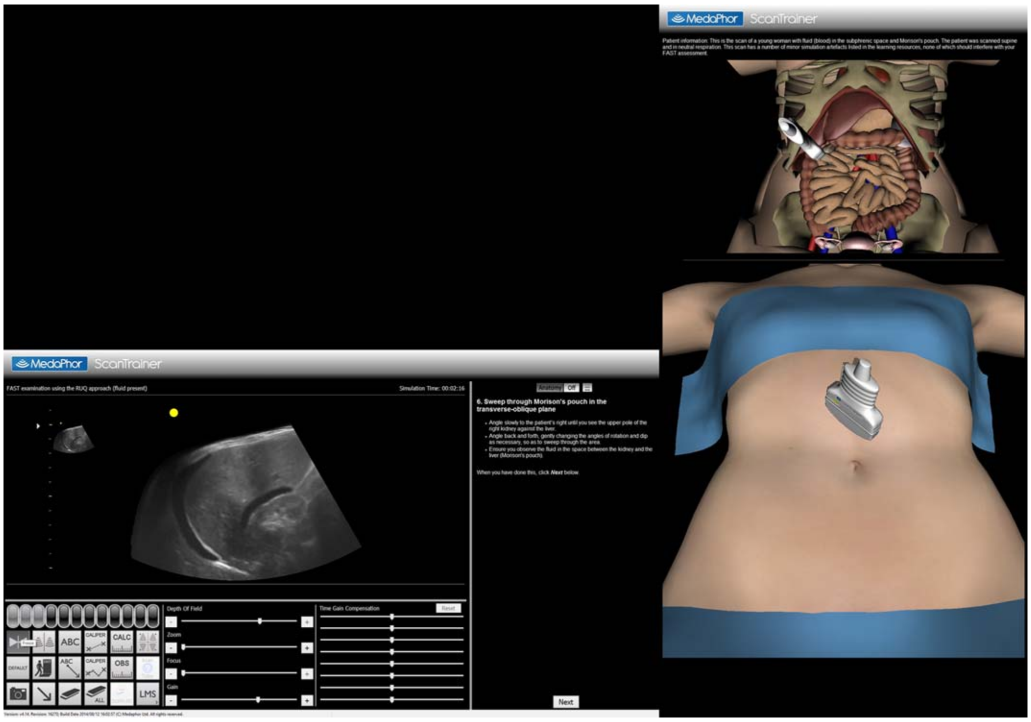
Step 5: Consequences

Test consequences are the expected and unexpected results of testing. Performance standards were established for the final simulator test by defining 2 performance levels. The first level was a pass/fail level, which was determined by the contrasting groups method.[21,23,24] By this method, a performance level was determined according to the score that allowed as few false-negative (failing experts) and false-positive (passing novices) results as possible. In practice, this level corresponds to the intersection between the distribution of scores in a group of competent performers (experts) and a group of noncompetent performers (novices). The second performance standard was determined according to the median score for the expert group (expert performance level).

Statistical Analysis

The participants' performances were measured by the built-in simulator metrics. These simulator metrics corresponded to either a pass or fail score. Each "passed" rating was assigned a score of 1, and each "failed" rating was assigned a score of 0. The Fisher exact test was used to determine whether there was a significant difference between the study groups for each metric. Simulator scores were calculated as percentages of the maximum simulator metric scores. Simulator score variance in the groups was assessed by the Levene test. The Mann-Whitney *U* test was used to compare simulator test scores. The internal consistency of all items in the final test was assessed by the Cronbach α statistic. Test-retest reliability was assessed by calculating the intraclass correlation coefficient between simulator scores on 2 attempts at the simulator test for the novice group. We chose a significance value of .01 to minimize the risk of inappropriate rejection of a true null hypothesis (type I error)

Figure 2. Screenshot of one of the FAST modules.



due to multiple comparisons.[25] Data analysis was conducted with SPSS version 23.0.0.0 software (IBM Corporation, Armonk, NY).

Results

A total of 10 ultrasound experts and 25 ultrasound novices were included. Participant characteristics are shown in Table 1. Of the initial 55 metrics, 34 (61.8%) had validity evidence ($P < .01$), and metrics from all 5 training modules were represented (Appendix 2). These metrics assessed various skills relating to the FAST examination, including image optimization/equipment handling (59%), systematic scanning technique (23%), image interpretation (15%), and speed of the examination (3%; Table 2). Of the 21 metrics that did not have validity evidence, 57% were related to image optimization/equipment handling; 33% were related to systematic scanning technique; and 10% were related to image interpretation. The distribution of metrics did not differ significantly between metrics with and without validity evidence ($P = .87$).

The novice group had a median score of 14.7% (range, 0%–47.1%), and the expert group had a median score of 94.1% (range, 94.1%–100%) on the simulator test ($P < .001$). The internal consistency of the 34 metrics, which constituted the simulator test, was high (Cronbach $\alpha = 0.98$). The test/retest reliability was high (intraclass correlation coefficient = 0.89).

The pass/fail level was determined by the contrasting groups method to be 79.7% of the maximum score (Figure 3). This level allowed no passing novices (false-positive) and no failing experts (false-negative). As illustrated in Figure 3, the expert group demonstrated less variance in their scores compared to the novice group (Levene test, $P = .001$). The expert performance level was determined from the median score of the expert group, which was 94.1%.

Table 1. Participant Demographics

Characteristic	Novices (n = 25)	Experts (n = 10)
Median age (range), y	27 (25–34)	53 (36–62)
Women, n (%)	12 (52)	3 (30)
Men, n (%)	13 (48)	7 (70)
Median experience (range), y	0	15 (2.5–30)

Table 2. Distribution of Metrics

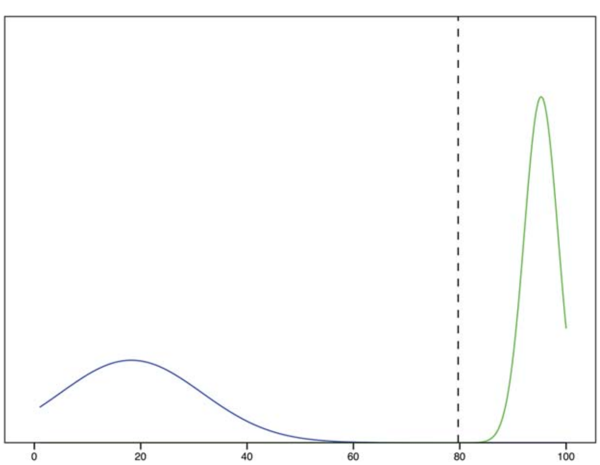
Metric	Valid	Nonvalid
Image optimization/equipment handling, n (%)[a]	20 (59)	12 (57)
Systematic scanning technique, n (%)[b]	8 (23)	7 (33)
Image interpretation, n (%)[c]	5 (15)	2 (10)
Speed of examination, n (%)[d]	1 (3)	0

[a]Organ/area correctly centralized, transducer placed in the correct location, transducer oriented in the correct plane, and transducer remained in the correct plane.
[b]Organ/area correctly visualized, and organ/area correctly examined.
[c]Borders marked correctly.
[d]Scan completed in appropriate time.

Discussion

This study demonstrates that FAST skills can be assessed in a reliable and valid way by using defensible performance standards on a virtual reality ultrasound simulator. Previous studies have demonstrated that not all metrics provided by the manufacturer discriminate between different levels of competence.[20,21] This study found that only 61.8% (34 of 55) of the metrics in this simulator test were able to discriminate between different levels of ultrasound experience, which underscores the need for a thorough evaluation of the validity of simulator metrics before using them for assessment purposes.

Figure 3. Pass/fail levels. The contrasting groups method was used to determine pass/fail levels of performances in the simulated setting. A pass/fail level was identified that ensured that as few competent operators as possible failed the simulator test (false-negative; green) and as few noncompetent operators as possible passed the test (false-positive; blue).



Simulation-based medical education enables training without patient risk or discomfort[26] and may accelerate learning curves as an adjunct to traditional training methods. However, as demonstrated in our study, there is a need for careful evaluation of how performance is assessed in the simulated setting. Without reliable and valid assessment scores and defensible performance standards, it is not possible to determine when trainees have reached their learning plateau or when they have achieved acceptable performance standards. In this study, we determined a pass/fail level that reliably discriminated between operators with different experience levels and resulted in no false-negative (failing competent operators) or false-positive (passing noncompetent operators) results. Although the pass/fail level discriminated well between operators with different levels of clinical experience, previous studies suggest that learning curves of novice operators may first plateau at expert levels of performance.[20,21] More studies are therefore needed to explore the learning curves for novice ultrasound operators in a simulated setting to determine how much training is required to surpass the pass/fail level and, in turn, to reach the expert performance level. The performance standards established in this study are to be considered milestones, which may be used to determine when novice operators are ready for clinical practice. However, the use of these milestones does not mean that novice operators who are trained according to these performance standards no longer need clinical training but, rather, that they have acquired basic skills before they begin practicing with patients.

Our study was limited by the fact that we only evaluated validity evidence of simulation-based assessments of FAST skills on one of the many commercially available systems. Hence, in terms of generalization of validity evidence, the specific scores derived from this study are not universally valid and should be reevaluated for different simulators, settings, and populations. In addition, we were only able to evaluate the 4 standard FAST windows on this simulator and not the extended examination recommended in the American Institute of Ultrasound in Medicine guideline for extended FAST.[4] However, the precautions noted regarding the indiscriminate use of simulator metrics are also relevant to other systems, and the approach to validity testing used in this study may be applied to other simulators. Another limitation was that the test-retest reliability was assessed by using the novice group's first 2 attempts at the simulator test. Because we may have expected some performance improvement from the first to the second attempt, the test-retest reliability demonstrated in our study may have represented a highly conservative estimate of the true reliability coefficient.

Finally, we chose to include participants who were either true experts (radiologists) or true novices (medical students). This approach allowed us to gain an insight into the learning potential that may be achieved in the simulated setting, although the use of either of these groups may not be representative of actual trauma team compositions. Future studies should include surgeons, emergency medicine physicians, and intensivists who perform FAST examinations with variable frequencies and should evaluate their performance on the basis of the objective standards established in this study.

Focused assessment with sonography for trauma skills can be assessed in a reliable and valid way against defensible performance standards by using virtual reality ultrasound simulators. Instead of requiring a fixed number of scans before commencement of independent practice, we recommend a competency-based approach in which trainee performance is assessed on the basis of objective criteria, such as those developed in this study.

Appendix 1: Training Modules and Tasks

Right Upper Quadrant

Task 1: Locate the right hemidiaphragm and subphrenic space in the sagittal plane.
Task 2: Sweep the subphrenic space in the sagittal plane.
Task 3: Locate the right hemidiaphragm in the transverse oblique plane.
Task 4: Sweep the subphrenic space in the transverse oblique plane.
Task 5: Outline the subphrenic fluid.
Task 6: Sweep through the Morison pouch in the transverse oblique plane.
Task 7: Outline the fluid in the Morison's pouch.

Left Upper Quadrant

Task 1: Locate the left hemidiaphragm and subphrenic space.
Task 2: Sweep the subphrenic space.
Task 3: Outline the subphrenic fluid.
Task 4: Sweep through the splenorenal space.
Task 5: Outline the fluid in the splenorenal space.

Subxiphoid Approach

Task 1: Locate the apex of the heart.

Task 2: Sweep through the apex of the heart in the transverse plane.

Task 3: Outline the pericardial fluid at the apex of the heart.

Pelvic Cavity

Task 1: Locate the pelvic cavity.

Task 2: Sweep the pelvic cavity in the sagittal plane.

Task 3: Outline the fluid in the pelvic cavity.

Task 4: Locate the bladder in the transverse plane.

Task 5: Sweep the pelvic cavity in the transverse plane.

Task 6: Outline the fluid in the pelvic cavity.

Final Examination

Task 1: Image the right subphrenic space.

Task 2: Image the right hepatorenal space.

Task 3: Image the pericardium.

Task 4: Image the left subphrenic space.

Task 5: Image the splenorenal space.

Task 6: Image the space around the pelvic cavity.

Task 7: Scan completed in appropriate time.

Appendix 2: Metrics That Had Validity Evidence (n = 34)

Right Upper Quadrant

Task 1: Locate the right hemidiaphragm and subphrenic space in the sagittal plane.

- Right upper quadrant 1.1: subphrenic space correctly centralized.
- Right upper quadrant 1.2: right hemidiaphragm visualized.
- Right upper quadrant 1.3: transducer placed in the correct location.
- Right upper quadrant 1.4: transducer orientated in the sagittal plane.

Task 3: Locate the right hemidiaphragm in the transverse oblique plane.

- Right upper quadrant 3.1: subphrenic space correctly centralized.
- Right upper quadrant 3.2: right hemidiaphragm visualized.
- Right upper quadrant 3.3: transducer placed in the correct location.
- Right upper quadrant 3.4: transducer orientated in the transverse oblique plane.

Task 5: Outline the subphrenic fluid.

- Right upper quadrant 5.1: borders of the subphrenic fluid marked correctly.

Task 7: Outline the fluid in the Morison pouch.

- Right upper quadrant 7.1: borders of the Morison pouch marked correctly.

Left Upper Quadrant

Task 1: Locate the left hemidiaphragm and subphrenic space.

- Left upper quadrant 1.1: subphrenic space correctly centralized.
- Left upper quadrant 1.2: pleural space visualized.
- Left upper quadrant 1.3: left hemidiaphragm visualized.
- Left upper quadrant 1.4: transducer placed in the correct location.
- Left upper quadrant 1.5: transducer orientated in the sagittal plane.

Task 3: Outline the subphrenic fluid.

- Left upper quadrant 3.1: borders of the subphrenic fluid marked correctly.

Task 4: Sweep through the splenorenal space.

- Left upper quadrant 4.1: splenorenal space correctly examined.
- Left upper quadrant 4.2: transducer remained in the correct plane.

Task 5: Outline the fluid in the splenorenal space.

- Left upper quadrant 5.1: borders of the splenorenal fluid correctly marked.

Subxiphoid Approach

Task 1: Locate the apex of the heart.

- Subxiphoid approach 1.1: apex of the heart correctly centralized.
- Subxiphoid approach 1.2: transducer placed in the correct location.
- Subxiphoid approach 1.3: transducer placed in the sagittal plane.

Pelvic Cavity

Task 1: Locate the pelvic cavity.

- Pelvic cavity 1.1: potential pelvic cavity correctly centralized.
- Pelvic cavity 1.2: uterus visualized.
- Pelvic cavity 1.3: rectum visualized.
- Pelvic cavity 1.4: bladder visualized.
- Pelvic cavity 1.5: transducer placed in the correct location.

- Pelvic cavity 1.6: transducer oriented in the sagittal plane.

Task 3: Outline the fluid in the pelvic cavity.

- Pelvic cavity 3.1: borders of the pelvic cavity marked correctly.

Task 4: Locate the bladder in the transverse plane.

- Pelvic cavity 4.1: bladder correctly centralized.

Final Examination

Task 4: Image the left subphrenic space.

- Final examination 4.1: left subphrenic space correctly centralized.

Task 5: Image the splenorenal space.

- Final examination 5.1: splenorenal space correctly centralized.

Task 6: Image the space around the pelvic cavity.

- Final examination 6.1: space around the pelvic cavity correctly centralized.

Task 7: Scan completed in appropriate time.

- Final examination 7.1: scan completed in an appropriate time.

References

1. Moore CL, Copel JA. Point-of-care ultrasonography. *N Engl J Med* 2011; 364:749–757.
2. American College of Surgeons Committee on Trauma. *Advanced Trauma Life Support Student Course Manual*. 9th ed. Chicago, IL: American College of Surgeons; 2012.
3. Osvalder J, Heinzmann A, Mathis G. E-FAST. In: Dietrich CF (ed). *EFSUMB Course Book on Ultrasound*. 1st ed. London, England: European Federation of Societies for Ultrasound in Medicine and Biology; 2013:1–30.
4. American Institute of Ultrasound in Medicine. AIUM practice guideline for the performance of the focused assessment with sonography for trauma (FAST) examination. *J Ultrasound Med* 2014; 33:2047–2056.
5. Tolsgaard MG, Ringsted C, Dreisler E, et al. Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial. *Ultrasound Obstet Gynecol* 2015; 46:312–318.
6. Shackford SR, Rogers FB, Osler TM, Trabulsy ME, Clauss DW, Vane DW. Focused abdominal sonogram for trauma: the learning curve of nonradiologist clinicians in detecting hemoperitoneum. *J Trauma* 1999; 46:553–562.
7. Jang T, Sineff S, Naunheim R, Aubin C. Residents should not independently perform focused abdominal sonography for trauma after 10 training examinations. *J Ultrasound Med* 2004; 23:793–797.
8. Ma OJ, Gaddis G, Norvell JG, Subramanian S. How fast is the focused assessment with sonography for trauma examination learning curve? *Emerg Med Australas* 2008; 20:32–37.
9. Tolsgaard MG, Todsen T, Sorensen JL, et al. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. *PLoS One* 2013; 8:e57687.
10. Griswold-Theodorson S, Ponnuru S, Dong C, Szyld D, Reed T, McGaghie WC. Beyond the simulation laboratory: a realist synthesis review of clinical outcomes of simulation-based mastery learning. *Acad Med* 2015; 90:1553–1560.
11. McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med* 2015; 90:1438–1441.
12. Knudson MM, Sisley AC. Training residents using simulation technology: experience with ultrasound for trauma. *J Trauma* 2000; 48: 659–665.
13. 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; 16:818–822.
14. Chaudery M, Clark J, Dafydd DA, et al. The face, content, and construct validity assessment of a focused assessment in sonography for trauma simulator. *J Surg Educ* 2015; 72:1032–1038.
15. Beaulieu Y, Laprise R, Drolet P, et al. Bedside ultrasound training using web-based e-learning and simulation early in the curriculum of residents. *Crit Ultrasound J* 2015; 7:1.
16. Salen P, O'Connor R, Passarello B, et al. FAST education: a comparison of teaching models for trauma sonography. *J Emerg Med* 2001; 20:421–425.
17. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005; 27: 10–28.
18. Mohammad A, Hefny AF, Abu-Zidan FM. Focused assessment sonography for trauma (FAST) training: a systematic review. *World J Surg* 2014; 38:1009–1018.
19. Blum T, Rieger A, Navab N, Friess H, Martignoni M. A review of computer-based simulators for ultrasound training. *Simul Healthc* 2013; 8:98–108.
20. Madsen ME, Konge L, Nørgaard LN, et al. Assessment of performance measures and learning curves for use of a virtual-reality ultrasound simulator in transvaginal ultrasound examination. *Ultrasound Obstet Gynecol* 2014; 44:693–699.
21. Dyre L, Nørgaard LN, Tabor A, et al. Collecting validity evidence for the assessment of mastery learning in simulation-based ultrasound training. *Ultraschall Med* 2016; 37:386–392.
22. Messick S. Validity. In: Linn RL (ed). *Educational Measurement*. New York, NY: MacMillan; 1989:13–104.
23. Tolsgaard MG, Ringsted C, Dreisler E, et al. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. *Ultrasound Obstet Gynecol* 2014; 43:437–443.

24. Livingston S, Zieky M. *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service; 1982.
25. Kirkwood BR, Sterne JA. Part B: analysis of numerical outcomes. Part C: analysis of binary outcomes. In: Goodgame F (ed). *Essential Medical Statistics*. London, England: Blackwell Science; 2003:31–223.
26. Tolsgaard MG, Ringsted C, Rosthøj S, et al. The effects of simulation-based transvaginal ultrasound training on quality and efficiency of care: a multicenter single-blind randomized trial. *Ann Surg* 2017; 265:630–637.